



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)	A2	(11) International Publication Number: WO 99/24578 (43) International Publication Date: 20 May 1999 (20.05.99)																					
(21) International Application Number: PCT/IB98/01665 (22) International Filing Date: 9 October 1998 (09.10.98) (30) Priority Data: <table border="0" style="width: 100%;"> <tr> <td style="width: 30%;">9723516.2</td> <td style="width: 40%;">6 November 1997 (06.11.97)</td> <td style="width: 30%;">GB</td> </tr> <tr> <td>9724190.5</td> <td>14 November 1997 (14.11.97)</td> <td>GB</td> </tr> <tr> <td>9724386.9</td> <td>18 November 1997 (18.11.97)</td> <td>GB</td> </tr> <tr> <td>9725158.1</td> <td>27 November 1997 (27.11.97)</td> <td>GB</td> </tr> <tr> <td>9726147.3</td> <td>10 December 1997 (10.12.97)</td> <td>GB</td> </tr> <tr> <td>9800759.4</td> <td>14 January 1998 (14.01.98)</td> <td>GB</td> </tr> <tr> <td>9819016.8</td> <td>1 September 1998 (01.09.98)</td> <td>GB</td> </tr> </table> (71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT). (72) Inventors; and (75) Inventors/Applicants (for US only): MASNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbucco, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).		9723516.2	6 November 1997 (06.11.97)	GB	9724190.5	14 November 1997 (14.11.97)	GB	9724386.9	18 November 1997 (18.11.97)	GB	9725158.1	27 November 1997 (27.11.97)	GB	9726147.3	10 December 1997 (10.12.97)	GB	9800759.4	14 January 1998 (14.01.98)	GB	9819016.8	1 September 1998 (01.09.98)	GB	(74) Agent: HALLYBONE, Huw, George; Carpmals & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
9723516.2	6 November 1997 (06.11.97)	GB																					
9724190.5	14 November 1997 (14.11.97)	GB																					
9724386.9	18 November 1997 (18.11.97)	GB																					
9725158.1	27 November 1997 (27.11.97)	GB																					
9726147.3	10 December 1997 (10.12.97)	GB																					
9800759.4	14 January 1998 (14.01.98)	GB																					
9819016.8	1 September 1998 (01.09.98)	GB																					
(54) Title: NEISSERIAL ANTIGENS (57) Abstract <p>The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.</p>																							

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

NEISSERIAL ANTIGENS

This invention relates to antigens from *Neisseria* bacteria.

BACKGROUND ART

Neisseria meningitidis and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (ie. having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

- 5 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide
10 sequences disclosed in the examples.

Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise
15 at least n consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, n is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences
20 complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as
25 those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (eg. expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines,
5 for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the
10 presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient
15 a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

20 A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

25 A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

5 General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and*
10 *ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene*
15 *Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

20 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

Definitions

25 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only

in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

5 A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing
10 sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

15 Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Bimstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In
20 *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

25 Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as
30 mammalian cells or bacteria. Mammalian replication systems include those derived from animal

- viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].
- 10 The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.
- 15 Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*eg.* Hep G2), and a number of other cell lines.
- 20 ii. Baculovirus Systems
- The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.
- 25
- After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques
- 30

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter

and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion
5 can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

10 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein,
15 which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from
20 wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

25 Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol.* 25:225).
30

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg. Summers and Smith supra.*

- 5 The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as
- 10 chromatography, *eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like.* As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, *eg. proteins, lipids and polysaccharides.*
- 15 In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

- 20 There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in
- 25 Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by
- 30 gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins: in: Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation
5 region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and
10 thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (E.
15 coli) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include
20 promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406]
25 promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac*
30 promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21].

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.*

(1989) *J. Gen. Microbiol.* 135:11], and Chey [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907],

Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology I*:412, *Streptococcus*].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence

of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples
5 include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

10 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters
15 which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA*
20 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

25 A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guillerimondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*];

[Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent
5 Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 Saccharomyces]; [Beach and Nurse (1981) *Nature* 300:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; Yarrowia].

Antibodies

10 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised
15 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably
20 a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection
25 is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating
30 the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is

recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described
5 above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of
10 the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then
15 cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes
20 are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A,
25 and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of
30 this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

5 Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or
10 antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine
15 experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such
20 as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus
25 particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack
30 Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,

picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

5 Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

10 Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

15 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

20 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

25 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654.

Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

- 5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317;
- 10 Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805;
- 15 Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240;
- 20 Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245;
- 25 Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre
- 30 (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

- 10 Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA*

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

15 Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

20 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hypodermic sprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

25 Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

- One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor, and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

- Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta*. 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to
5 mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1,2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand
10 Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

15 Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate
20 ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta*
25 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

- 5 Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can
10 be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful
15 as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the
20 list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

25 Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

“Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

- total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.
- Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

- In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

- In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology,

and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed
5 after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid
10 probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will
15 encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some
20 variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe
25 sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as
30 temperature, salt condition and the like. For example, for diagnostic applications, depending on the

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

- 5 Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.*
10 backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

- 15 Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with
20 duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern
25 blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid
30 support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

5 **Figures 1-20** show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* **143**:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* **12**:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl. 11:9) and is available in the Protean package of DNASTAR, Inc. 15 (1228 South Park Street, Madison, Wisconsin 53715 USA).

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

20 The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in 25 *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psort.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

10 In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8).
15 After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one ChCl_3 /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70%
20 ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A
25 sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included

a *XhoI* restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *BamHI-XhoI* or *EcoRI-XhoI*), and pET21b+ (using either *NdeI-XhoI* or *NheI-XhoI*).

5 5'-end primer tail: CGCGGATCCCATATG (*BamHI-NdeI*)
 CGCGGATCCGCTAGC (*BamHI-NheI*)
 CCGGAATTCTAGCTAGC (*EcoRI-NheI*)
 3'-end primer tail: CCCGCTCGAG (*XhoI*)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF;
 10 the same 3' *XhoI* primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (*NdeI*)
 5'-end primer tail: CGGGATCC (*BamHI*)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *NheI*-
 15 *BamHI* restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (*NheI*)
 3'-end primer tail: CGGGATCC (*BamHI*)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing
 20 nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad \text{(tail excluded)}$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad \text{(whole primer)}$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and
 25 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH₄OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100μl or 1ml of water. OD₂₆₀ was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/μl.

C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40μM of each oligo, 400-800μM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl₂), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10μl DMSO or 50μl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

	95°C	65-70°C	72°C
--	------	---------	------

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

- 5 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

10 D) Digestion of PCR fragments

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- 15
 - *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
 - For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
 - *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of
- 20 the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or

- 25 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the
10 vector pTRC99 (Pharmacia).

F) Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer.
15 The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then
20 centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin
25 Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SaII* or, for ORF 122, *SaII-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia)

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD₂₈₀ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

- 5 For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation
10 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM
15 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D₂₈₀ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D₂₈₀ was 0.1. 21µl of each
20 fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-
25 14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

10 M) Mice immunisations

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)₃ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)₃, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

20 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA was considered positive when OD₄₉₀ was 2.5 times the respective pre-immune sera.

10 O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

15 R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₆₂₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₆₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and purification results.

Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACCTGC
101 A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501 AGACCG...

```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAAQGN AAAQYNLGAM
51  YXQRTVRVRD DAEAVRWYRQ PAEQGLAQAO YNLGWMYANG RXVRQDDTEA
101 VRWYRQAAAO GVYQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
151 AQNNLGVMYA ERXRVQRD...

```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACCTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501 AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
551 ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA

```

5

1 MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
51 YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAO YNLGWMYANG RGVRQDDTEA
101 VRWYRQAAAO GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAOGVAQ
151 AQNNLGVMYA ERRGVRQDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY*

10

```

1  ATGAAACAGA  CAGTCAAATG  GCTTGCCGCC  GCCCTGATTG  CTTGGGCTT
51  GAACCAAGCG  GTGTGGGCGG  ATGACGTATC  GGATTTTCGG  GAAACTTGC
101 AGGCGGCAGC  ACAGGGAAAT  GCAGCAGCCC  AAAACAATTT  GGGCGTGATG
151 TATGCCGAAA  GACGCGGCGT  GCGCCAAGAC  CGCGCCCTTG  CACAAGATG
201 GCTTGGCAAG  GCTTGTCAA  ACGGATACCA  AGACAGCTGC  GACAATGACC
251 AACTGCCTGAA  AGCGGGTTAT  TGA

```

15

1 MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
51 YAERRGVROD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *

20

```

              10      20      30      40      50      60
orf37.pep    MKQTVXMLAAALIALGLNRPVWXXDDVSDFRENLXAAAQGNAQAQYNLGAMYXQRTVRRD
             |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf37a       MKQTVKWLAAALIALGLNQAVWADDVSDFRENLQAAAQNNGVMYAERRGVQRD
              10      20      30      40      50      60

```

25

```

              70          80          90          100          110          120
orf37.pep    DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
              ||| :| : :||
orf37a       RALAQEWLGKACQNGYQDSCDNDQRLKAGYX
              70          80          90

```

30

35

1	ATGAAACAGA	CAGTCAAATG	GCTTGCCGCC	GCCCTGATTG	CCTTGGGCTT
51	GAACCAAGCG	GTGTGGGCGG	GTGACGTATC	GGATTTTCGG	GAAAACATTG
101	AGGcggcaga	ACagggAAAT	GCAGCAGCCC	AATTCAATTT	GGCGCTGATG
151	TATGAAATG	GACAAGGAGT	TCGTCAAGAT	TATGTACAGG	CAGTCGATG
201	GTATCGCAAG	GCTTCAGAAC	AAGGGGATGC	CCAAGCCCAA	TACAATTTGG
251	GCTTGATGTA	TTACGATGGA	CGCGGCTGTC	GCCAAGACCT	TGCGCTCGCT
301	CAACAATGGC	TTGGCAAGGC	TTGTCAAAC	GGAGACCAA	ACAGTCGCA
351	CAATGACCAA	CGCCTGAAGG	CGGGTTATTA	A	

40

1 MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
51 YENGQGVQRD YVQAVQWYRK ASEQGDAAQ YNLGLMYDYG RGVQRDLALA
101 QOWLKGACON GDQNSCDNDQ RLKAGY*

45

```

orf37.pep      MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAAQGNAAQYNLGAMYXQRTVRRRD      60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf37na       MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAQFNLGVMYENGQGVRRD      60

```

50

```

orf37.pep      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG 120
               ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
orf37ng       YVQAVQWYRKASEQGDAAQAYNLGLMYDDGRGVRQDLALAAQQWLKGACQNGDQNSCDNDQ 120

```

orf37.pep VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERXRVQRD 168

55

orf37ng RLKAGY 126

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

```

      10      20      30      40      50      60
orf37-1.pep MKQTVKWLAAALIALGLNRAVWADDVSDFFRENLQAAAQGNAAQYNLGAMYYKGRGVRRD
5 orf37ng    MKQTVKWLAAALIALGLNQAVWAGDVSDFFRENLQAAEQGNAAQFNLGVMYENGQGVRRD
      10      20      30      40      50      60

      70      80      90     100     110     120
orf37-1.pep DAEAVRWYRQAAEQGLAQYNLGWMYANGRGVRQDDTEAVRWYRQAAAQGVVQAQYNLG
10 orf37ng    YVQAVQWYRKASEQGDAAQYNLGLMYDGRGVRRD-----
      70      80      90

      130     140     150     160     170     180
orf37-1.pep VIYAEGRGVRQDDVEAVRWFRQAAAQGVAAQNNLGVMYAERRGVRRQDRALAEWFGKAC
15 orf37ng    -----LALAQQWLKAC
                                100

      190     199
orf37-1.pep QNGDQDGCNDQRLKAGYX
20 orf37ng    QNGDQNSCNDQRLKAGYX
      110     120

```

- 25 Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 30 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

- 35 Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

```

      TTCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA
40 GGGGTATTGG TCGGGCGCGT CGGCGCTATC GGAATTGACC CGAAATCCTA
      TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG
      ACGTTTCCGC GCAAATCCTG ACTTCSGGAC TTTTGGGCGA GCAGTACATC
      GGGCTGCAGC AGGGCGGCGA CACGGAAAAC CTTGCTGCCG GCGACACCAT
      CTCCGTAACC AGTTCTGCAA TGGTCTGGA AAACCTTATC GGCAAATTCA
45 TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGGCAATGC GGAAAAAGCC
      GCCGAATAA

```

This corresponds to the amino acid sequence <SEQ ID 10>:

```

      1 FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD
51 VSAQILTSLG LGEQYIGLQQ GGDTENLAAG DTISVTSSAM VLENLIGKFM

```

101 TSFAEKNADG GNAEKAAE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

```

5      20      30      40      50      60      70
ybrd.h LGIGALVFLGLRVANVQGF AETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
              10      20      30

10     80     90     100     110     120     130
ybrd.h KSYLPKVSIAINQ EYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGS QIQDT
N.m      KSYQARVRLDLGKY-QFSSDVSAQILTSGLLGEQYIGLQQG---GDTENLAAGDTISVT
              40      50      60      70      80

15     140     150     160
ybrd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
N.m      SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX
              90      100     110     120

```

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

```

25     20     30     40     50     60     70
ybrd      GAAAVAFLAFRVAGGA FGGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
              10     20     30

30     80     90     100     110     120     130
ybrd      KSYQARVRLDLGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
N.m      KSYQARVRLDLGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
              40     50     60     70     80     90

35     140     150     160
ybrd      VLENLIGKFMTSFAEKN AEGGNAEKAAEX
N.m      VLENLIGKFMTSFAEKNADGGNAEKAAEX
              100     110     120

```

The complete ybrd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```

50      1  ..ATTTTGATAT ACCTCATCCG CAAGAATCTA GGTTCCGCCG TCTTCTTCTT
      51  TCAGGAACGC CCCGGAAAGG ACGGAAAACC TTTTAAATG GTCAAATTC
      101 GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
      151 GAACGCCTGA CACCGTTCGG CAAAAAAGT CGTGCCGcCA GTwTGGACGA
      201 ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
      251 CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAC
      301 CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG

```

-64-

5 351 GCGCAACGCG CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
 401 TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
 451 AAAAAAGTAT TAATCAAGGA AGGGATTTC GCACAGGGCG AACA.aCCAT
 501 GCCCCCTTTC ACAGGAAAAC GCAAACTCGC CGTCGTCGGT GCGGGCGGAC
 551 ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA
 601 ATCGTTTTTC TGGACGACCG CGCACAAAGG AGCGTCAACG GCTTTTCCGT
 651 CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
 701 ACGTCGCCGT CGCCGTCGGC AACAAACCGCA TCCGCCGCCA AATCGCCGAA
 751 AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC
 10 801 GACCGTCTCG CCTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
 851 AAGCGGTCG.

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

15 1 . . I L I Y L I R K N L G S P V F F F Q E R P G K D G K P F K M V K F R S M R D G L Y S D G I P L P D G
 51 E R L T P F G K K L R A A S X D E L P E L W N I L K G E M S L V G P R P L L M Q Y L P L Y D N F Q N
 101 R R H E M K P G I T G W A Q V N G R N A L S W D E K F A C D V W Y I D H F S L C L D I K I L L L T V
 151 K K V L I K E G I S A Q G E X T M P P F T G K R K L A V V G A G G H G K V V A D L A A L G R Y R E
 201 I V F L D D R A Q G S V N G F S V I G T T L L E N S L S P E Q Y D V A V A V G N N R I R R Q I A E
 251 K A A A L G F A L P V L V H P D A T V S P S A T V G Q G S V V M A K A V . .

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

20 1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
 51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
 101 AGAATCTAGG TTCGCCGTC TTCTTCTTTC AGGAACGCC CCGAAAGGAC
 151 GGAAACCTT TTAATAGGT CAAATCCGT TCCATGCGCG ACGCGCTTGA
 201 TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
 25 251 AAAAATGCG TGCCGCCAGT TTGGACGAAC TGCCTGAATT ATGGAATATC
 301 TTAAGAGCG AGATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
 451 GAAAATTTCG CTTGCGATGT TTGGTATATC GACCACTCA GCCTGTGCCT
 30 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
 551 GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
 601 AAATCGCCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT
 651 TGCCGCCGCA CTCGCCGCT ACAGGGAAT CGTTTTTCTG GACGACCGCG
 701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
 35 751 GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
 851 CCCTGCCCGT TCTGTTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
 901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
 951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCAGC
 40 1001 ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCTG
 1051 GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
 1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
 1201 AAGCCGCTGC CGCGCAAAA CCCCAGAGACC TCGACAGCAT AA

45 This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

50 1 M S K F F K R L F D I V A S A G L I F L S P V F L I L I Y L I R K N L G S P V F F F Q E R P G K D
 51 G K P F K M V K F R S M R D A L D S D G I P L P D G E R L T P F G K K L R A A S L D E L P E L W N I
 101 L K G E M S L V G P R P L L M Q Y L P L Y D N F Q N R R H E M K P G I T G W A Q V N G R N A L S W D
 151 E K F A C D V W Y I D H F S L C L D I K I L L L T V K K V L I K E G I S A Q G E A T M P P F T G K R
 201 K L A V V G A G G H G K V V A D L A A L G R Y R E I V F L D D R A Q G S V N G F S V I G T T L L L
 251 E N S L S P E Q Y D V A V A V G N N R I R R Q I A E K A A A L G F A L P V L V H P D A T V S P S A T
 301 V G Q G S V V M A K A V V Q A G S V L K D G V I V N T A A T V D H D C L L N A F V H I S P G A H L S
 351 G N T H I G E E S W I G T G A C S R Q Q I R I G S R A T I G A G A V V V R D V S D G M T V A G N P A
 401 K P L P R K N P E T S T A *

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:

10

20

30

-65-

```

orf3.pep                               ILIYLRKNLGSVPVFFQERPGKDGKPFKMVKFR
orf3a      MSKFFKRLFDIVASASGLIFLSPVFLILIYLRKNLGSVPVFFQERPGKDGKPFKMVKFR
           10      20      30      40      50      60

5      40      50      60      70      80      90
orf3.pep  SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL
orf3a      SMHDALDSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL
           70      80      90      100     110     120

10      100     110     120     130     140     150
orf3.pep  YDNFQNRHRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCLDIKILLTVKKVL
orf3a      YDNFQNRHRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSLCLDIKILLTVKKVL
           130     140     150     160     170     180

15      160     170     180     190     200     210
orf3.pep  IKEGISAQGEEXTMPFFTGRKRLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG
orf3a      IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRVQGSVNG
           190     200     210     220     230     240

20      220     230     240     250     260     270
orf3.pep  FSVIGTTLLENLSLSPSEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT
orf3a      FSVIGTTLLENLSLSPSEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLVHPDSTVSPSAT
           250     260     270     280     290     300

30      280
orf3.pep  VGQGSVVMKAV
orf3a      VGQGGVVMKAVVQADSVLKDGIVIVNTAATVDHDCLLDAFVHISPGAHLNTRIGEEESW
           310     320     330     340     350     360

```

35 The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

```

1  ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
101 AGAATCTGGG TTCGCCGCTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
151 GGAAAACCTT TTAATATGGT CAAATTCCTG TCCATGCACG ACGCGCTTGA
40 201 TTCAGACGGC ATTCTGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
251 AAAAATGCG TGCCGCCAGT TTGGACGAAC TGCCCGAACT GTGGAACGTC
301 CTCAAAGGCG ACATGAGCCT GGTCGGCCCC CGCCCGCTGC TGATGCAATA
351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCGG
401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
45 451 GAACGCTTCG CATGCGACAT CTGGTATATC GACCACTTCA GCGTGTGCTT
501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAAGAAG
551 GGATTCCGCG ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
601 AAACCTGCGC TCCTCGGTGC GGGCGGACAC GGCAAAGTCG TTGCCGAGCT
651 TGCCGCGGCA CTGCGACAT ACGGCGAAAT CGTTTTTCTG GACGACGCG
50 701 TCCAAGGCAG CGTCAACGGC TTCCCCGTCA TCGGCACGAC GCTGCTGCTT
751 GAAAACAGTT TATCGCCCGA ACAATTCGAC ATCGCCGTCG CCGTCGGCAA
801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
851 CCCTGCCGCT CCTGATTCAT CCGGACTCGA CCGTCTCGCC TTCTGCAACA
901 GTCGGACAAG GCGGCGTCGT TATGGCGAAA GCCGTGCTAC AGGCTGACAG
55 951 CGTATTGAAA GACGGCGTAA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001 ATTGCCTGCT TGATGCTTTC GTCCACATCA GCCCGGGCGC GCACCTGTCT
1051 GGCAACACGC GTATCGGCGA AGAAAGCTGG ATAGGCACAG GCGCGTGCAG
1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
1151 TCGTCGTGCG CGACGTTTCA GACGCGATGA CCGTCGCGGG CAACCCGGCA
60 1201 AAACCATTGG CAGGCAAAAA TACCAGACC CTGCGGTCGT AA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

```

1  MSKFFKRLFD IVASASGLIF LSPVFLILIY LIRKNLGSVP FFFQERPGKD
51 GKPFKMVKFR SMHDALDSDG ILLPDGERLT PFGKKLRAAS LDELPELWNV
101 LKGDSL VGPRPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD
65 151 ERFACDIWYI DHFSLCLDIK ILLTVKKVL IKEGISAQGE ATMPFFTGRK
201 KLAVVGAGGH GKVVAEELAAA LGTYGEIVFL DDRVQGSVNG FSVIGTLLL
251 ENSLSPEQFD IAVAVGNNRI RRQIAEKAAA LGFALPVLH PDSTVSPSAT

```

301 VGQGGVVMK AVVQADSVLK DGIVVNTAAT VDHDCLLDAF VHISPGAHL
 351 GNTRIGEE SW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
 401 KPLAGKNTET LRS*

Two transmembrane domains are underlined.

5 ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

		10	20	30	40	50	60
	orf3a.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKNLGSPVFFQERPGKDGKPFKMKVFR					
	orf3-1	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKNLGSPVFFQERPGKDGKPFKMKVFR					
10		10	20	30	40	50	60
	orf3a.pep	SMHDAALSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL					
	orf3-1	SMRDAALSDGIPLPDGERLTPFGKKLRAASLDELPELWNILKGEMLSVGPRPLLMQYLPL					
15		70	80	90	100	110	120
	orf3a.pep	YDNFQNRHHEMKPGITGWAQVNGRNLSDWERFACDIWYIDHFSCLLDIKILLTVKKVL					
	orf3-1	YDNFQNRHHEMKPGITGWAQVNGRNLSDWEKFAVDWYIDHFSCLLDIKILLTVKKVL					
20		130	140	150	160	170	180
	orf3a.pep	IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKVVAAELAAALGTGEIVFLDDRVQGSVNG					
	orf3-1	IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG					
25		190	200	210	220	230	240
	orf3a.pep	FPVIGTTLLENLSLSPQFDIAVAVGNNRIRRQIAEKAAALGFALPVLHPDSTVSPSAT					
	orf3-1	FSVIGTTLLENLSLSPQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
30		250	260	270	280	290	300
	orf3a.pep	VGQGGVVMKAVVQADSVLKDGIVVNTAATVDHDCLLDAFVHISPGAHLGNTRIGEE SW					
	orf3-1	VGQGSVVMKAVVQAGSVLKDGIVVNTAATVDHDCLLNAFVHISPGAHLGNTHIGEE SW					
35		310	320	330	340	350	360
	orf3a.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLAGKNTETLRSX					
	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
40		370	380	390	400	410	
	orf3a.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					

Homology with hypothetical protein encoded by yvfc gene (accession Z71928) of *B. subtilis*

ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

50	ORF3	3	IYLIRKNLGSPVFFQERPGKDGKPFKMKVFRSMRDGLYSDGIPLPDGERLTPFGKKLRA	62
	yvfc	27	IYVRLKIGSPVFFKQVRPGLHGKPFETLYKFRMTDERDSKGNLLPDEVRLTKTGRLIRK	86
55	ORF3	63	ASXDELPPELWNILKGEMLSVGPRPLLMQYLPLYDNFQNRHHEMKPGITGWAQVNGRNL	122
	yvfc	87	LSIDELPQLNLVNLKGDLSLVGPRPLMDYLPLYTEKQARRHEVKPGITGWAQINGRNAIS	146
60	ORF3	123	WDEKFAVDWYIDHFSCLLDXXXXXXXXXXXXXXXXXEGISAQEXTMPFFT	172
	yvfc	147	WEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEGIQQTNHVTAERFTG	196

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N. gonorrhoeae*:

5	orf3	ILIIYLIRKNLGSPVFFFQERPDKGPKFKMVKFR	34
	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKNKGSVFFIRERPGDKGPKFKMVKFR	60
10	orf3	SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL	94
	orf3ng	SMRDALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL	120
15	orf3	YDNFQNRHHEMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLCDIKILLTIVKKVL	154
	orf3ng	YNKFQNRHHEMKPGITGWAQVNGRNALSWDEKFSADVWYTDNFSFVLDKILFLTVKKVL	180
20	orf3	IKEGISAQGEATMPFFAGNRKLAIVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng	IKEGISAQGEATMPFFAGNRKLAIVGAGGHGKVVADLAAALGRYREIVFLDDRTQGSVNG	240
25	orf3	FSVIGTTLLENSLSPEQYDVAVAVGNRRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	274
	orf3ng	FPVIGTTLLENSLSPEQFDITVAVAVGNRRIRRQITENAAALGFALPVLVHPDATVSPSAI	300
30	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLGSGNTRIGEEESR	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

1	ATGAGTAAAG	CCGTCAAACG	CCTGTTTCGAC	ATCATCGCAT	CCGCATCGGG
51	GCTGATTGTC	CTGTCGCCCC	TGTTTTTGGT	TTTAATATAC	CTCATCCGCA
101	AAAACCTTAGG	TTCCGCCGTC	TTCTTCattC	GGGAACGCCc	cgGAAAGGAc
151	ggaaaacCTT	TTAAATGGT	CAAATTCCTG	TCCAtgcgcg	acgcgcttGA
201	TTCAGACGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTCCGGCA
251	AAAAATTACG	CGCCACCACT	TTGGACGAAC	TTCTGAATT	ATGGAATGTC
301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
351	TCTGCCGCTT	TACAACAAAT	TTCAAACCG	CGCCACGAA	ATGAAACCGG
401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
451	GGAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTC	GCTTTTGGCT
501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
551	GCATTTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
701	CCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
751	GAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCACCCTCG	CCGTCGGCAA
801	CAACCCGATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
851	AACTGCCCGT	TCTGATTCAT	CCCGACGCGA	CCGTCTCGCC	TTCTGCAATA
901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCTGTAC	AGGCCGGCAG
951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCCGGGCG	GCACCTGTCT
1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgcccGT	GCAGGGgcGG
1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTCTCGGG	CAACCCGGCA
1201	AAGCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence <SEQ ID 18>:

1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKNLGSPV	FFIRERPGKD
55	51	GKPFKMKVFR	SMRDALDSDG	IPLPDSERLT	DFGKKLRATS
101	101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRHE	MKPGITGWAQ
151	151	EKFSCDVWYT	DNFSFVLDK	ILFLTVKKVL	IKEGISAQGE
201	201	KLAVIGAGGH	GKVVAEELAA	LGTYGEIVFL	DDRTQGSVNG
251	251	ENSLSPEQFD	ITVAVGNRI	RRQITENAAA	LGFKLPVLIH
301	301	IGQGSVVMKAV	AVVQAGSVLK	DGVIVNTAAT	VDHDCLLDAF
351	351	GNTRIGEEESR	IGTGACSRQQ	TTVGSGVTAG	AGAVIVCDIP
401	401	KPLTGKNPKT	GTA*		DGMTVAGNPA

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

		10	20	30	40	50	60
	orf3-1.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNLGSPVFFQERPGKDGPVKFMVKFR					
5	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLRKKNLGSPVFFIRERPGKDGPVKFMVKFR					
		10	20	30	40	50	60
	orf3-1.pep	SMRDALDSGDIPLDGERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLP					
10	orf3ng	SMRDALDSGDIPLDGERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLP					
		70	80	90	100	110	120
	orf3-1.pep	YDNFQNRHHEMKPGITGWAQVNGRNLASWDEKFACDVWYIDHFSCLDIKILLTVKKVL					
15	orf3ng	YNKFQNRHHEMKPGITGWAQVNGRNLASWDEKFCSDVWYTDNFSFWLDMKILFLT					
		130	140	150	160	170	180
	orf3-1.pep	IKEGISAQGEATMPFFTGRKLAUVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG					
20	orf3ng	IKEGISAQGEATMPFFAGNRKLAUVGAGGHGKVVAAELAAALGTGEIVFLDDRTQGSVNG					
		190	200	210	220	230	240
	orf3-1.pep	FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
25	orf3ng	FPVIGTTLLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFKLPVLIHPDATVSPSAI					
		250	260	270	280	290	300
	orf3-1.pep	VGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEEESW					
30	orf3ng	IGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEEESR					
		310	320	330	340	350	360
	orf3-1.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
35	orf3ng	IGTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTX					
		370	380	390	400	410	
	orf3-1.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
40	orf3ng	IGTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTX					
		370	380	390	400	410	

In addition, ORF3ng shows significant homology with a hypothetical protein from *B.subtilis*:

45	gnl PID e238668 (271928) hypothetical protein [Bacillus subtilis]
	>gi 1945702 gnl PID e313004 (294043) hypothetical protein [Bacillus subtilis]
	>gi 2635938 gnl PID e1186113 (299121) similar to capsular polysaccharide biosynthesis [Bacillus subtilis] length = 202
	Score = 235 bits (594), Expect = 3e-61
	Identities = 114/195 (58%), Positives = 142/195 (72%)
50	Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLRKKNLGSPVFFIRERPGKDGPVKFMVKFRSMRD 64
	+KRLFD+ A+ L S + L I ++R +GSPVFF + RPG GKPF + KFR+M D
	Sbjct: 3 LKRLFDLTAAIFLLCCTSVIILFTIAVRLKIGSPVFFKQVRPGLHGKPFPLYKERTMTD 62
55	Query: 65 ALDSGDIPLDGERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPYLNKF 124
	DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRPLLM YLPLY +
	Sbjct: 63 ERDSKGNLLPDEVRLTKTGRLIRKLSIDELPQLLNVLKGDLSLVGPRPLLM DYLPYLYTEK 122
60	Query: 125 QNRHHEMKPGITGWAQVNGRNLASWDEKFCSDVWYTDNFSFWLDMKILFLT
	Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG
	Sbjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEG 182
	Query: 185 ISAQGEATMPFFAGN 199
	I T F G+
65	Sbjct: 183 IQQTNHVTAEERFTGS 197

The hypothetical product of *yvf* gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

```

1  ..AACCATATGG CGATTGTCAT CGACGAATAC GGCGGCACAT CCGGCTTGTT
51  CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT
101 TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301 TTTGCAGTTC ACCGTGCGAC GCGCCGACAA CCGCCGCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA
401 CAGTTTAG

```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51  WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTSR RARRKSPYRR
101 FAVHRRTRRQ PPPAYADGDP REV.....XR RCTV*

```

20 Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```

1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCG
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTTCGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251 CCTACGTTAT CGATACGCC CATTCGCGCT TCCCGTCAT CCGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAGA GTTCCGCGAA
451 CAGCGCAACC ATATGGCGAT TGTCATCGAC GAATACGGCG GCACATCCGG
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
651 CTTCGGCAGC GAATACAGCA GCGAAGAAGC CGACACCAT CCGCCTGGTC
701 ATTCAGAGT TGGGACATCT GCGCGTGGC GCGGAAAAAG TCCTTATCGG
751 CGGTTTGCAG TTCACCGTC CACGCGCCGA CAACCGCCGC CTGCATACCG
801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
851 TGACGGTACG GCGGTTTTCT GTTCAATCC GCCCATCCG CCAAACATAA

```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```

40 1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLLRLE
51  KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEEADTI RPHSRVGTSA ARARRKSPYR
45 251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCG
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTTAT CGATACGCC CATTCGCGCT TCCCGTCAT CCGTGAAGAC
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCCGTCT

```

15

20 The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa overlap with ORF5a:

```

25 orf5.pep                               10      20      30
                                         NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI
                                         |||||
orf5a      FHLKSILRPVAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI
              130      140      150      160      170      180

30 orf5.pep              40      50      60      70      80      90
                                         EDEFDEDDSadNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA
                                         |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf5a      EDEFDEDSADNIHAVSAERWRIHAATEIEDINAFGTEYSSEEADTIGGXGHSGIGTPA
              190      200      210      220      230      240

35 orf5.pep              100      110      120      130
                                         RARRKSPYRRFAVHRRTTRQPPPAYADGDPREVSVXXXXXXRRFCTV
                                         ||||| ||| | | | : |||||
orf5a      RARRKSXYRRXAXHXRXXQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXTX
              250      260      270      280      290      300

```

The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:

40		10	20	30	40	50	60
	orf5a.pep	MDGAQPKTNFXXR	LIARLAREPDS	AEDVLTLLRQA	HEQEVFDADT	LLRLEKVLDF	SDLEV
	orf5-1	MDGAQPKTNFFER	LIARLAREPDS	AEDVLNLLRQA	HEQEVFDADT	LLRLEKVLDF	SDLEV
		10	20	30	40	50	60
45		70	80	90	100	110	120
	orf5a.pep	RDAMITRSRMNVL	KENDSIERITAY	VIDTAHSRFPV	IGEDKDDEV	LGIHAKDLLK	YMFNP
	orf5-1	RDAMITRSRMNVL	KENDSIERITAY	VIDTAHSRFPV	IGEDKDDEV	LGIHAKDLLK	YMFNP
		70	80	90	100	110	120
50		130	140	150	160	170	180
	orf5a.pep	EQFHLKSILRPA	VFVPEGKSLT	ALLKEFREQR	NHMAIVIDEY	GGTSGLVT	FEDIIEQIVG
	orf5-1	EQFHLKSILRPA	VFVPEGKSLT	ALLKEFREQR	NHMAIVIDEY	GGTSGLVT	FEDIIEQIVG
		130	140	150	160	170	180
55		190	200	210	220	230	240
	orf5a.pep	DIEDEFDEDES	ADNIHAVSAER	WRIHAATEIE	DINAFFGTEY	SSEADTIGG	XGHSGIGT
		:	:	:	:	:	:
	orf5-1	EIEDEFDEDD	SADNIHAVSS	ERWRIHAATE	IEDINTFFGTEY	SSEADTIRP	-GHSRVGT
		190	200	210	220	230	
60		250	260	270	280	290	300

5 Further work identified the a partial DNA sequence in *N.gonorrhoeae* <SEQ ID 25> which encodes
a protein having amino acid sequence <SEQ ID 26; ORF5ng>:

Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:

This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:

The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng):

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in

orf5nq-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFAELEV

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

ORF5 and TlyC proteins show 58% aa identity in 77 aa overlap (BLASTp).

ORF5ng-1 also shows significant homology with TlyC:

BNSDOCID: <WO 9924578A2_1_>

		130	140	150	160	170	180
5	orf5ng-1.pep	170	180	190	200	210	220
		VT	FED	IIEQ	IVGD	IEDEF	DEDES
	tlyc_haein	VT	IED	IIEQ	IVGD	IEDEF	DEEEEIAD
		190	200	210	220	230	
10	orf5ng-1.pep	230	240	250	260	270	280
		TIR	RLGH	SGIG	-TPAR	ARRK	SPYRR
	tlyc_haein	TIG	GLIM	QTFG	YLPK	RGEI	IILKN
		240	250	260	270	280	290

15 Homology with a hypothetical secreted protein from *E.coli*:

ORF5a shows homology to a hypothetical secreted protein from *E.coli*:

20 sp|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION
>gi|1778577 (U82598) similar to *H. influenzae* [*Escherichia coli*] >gi|1786879
(AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
approx. 440 aa protein YTFL_HAEIN SW: P44717 [*Escherichia coli*] Length = 292

Score = 212 bits (533), Expect = 3e-54

Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

25 Query: 2 DGAQPKTNFXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60
D K F L+++L EP + +++L L+R + + ++ D DT LE V+D +D V
Sbjct: 10 DTISNKKGFFSLLLSQLFHGEPKNRDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

30 Query: 61 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
RD MI RS+M LK N +++ +I++AHSRFPVI EDKD + GIL AKDLL +M +
Sbjct: 70 RDIMIPRSQMITLKRNQTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

35 Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIV 179
E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
Sbjct: 130 AEAFSMDKVLRLQAVVPESKRVDRMLKEFRSQRVHMAIVIDEFGGVSGLVTIEDILELIV 189

Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEDT 229
G+IEDE+DE++ D +S W + A IED N FGT +S EE DT
Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVD 238

40 Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from
H. influenzae (hemolysins are secreted proteins), it was predicted that the proteins from
N.meningitidis and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or
diagnostics.

45 ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The
products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows
the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used
to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments
confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 5

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

1 ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTG
51 GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCC GACATCGGAC

101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
 151 GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTCCCGG ACAGCTACGA
 201 AATCGATGCG GCGGCGAGTG ATTTGCAGAT TTACCAAACC GCCTACAAGG
 251 GCGATGCAAC GCCGCCTGAA TGAGGGCATG GGAAAGCAGG CAGGACGGGC
 5 301 TGCCTTATAA AAACCCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
 351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
 401 CAACCGCCTG AAAATCGGTA TCGGCCTGCA AACCgAssCG TCCGTGATTT
 451 ACGGCATGGG TCGGCGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
 501 CGCGACACGC CGTACAACAC CTACACGCGC GCGGTCTGCG CGCCAACCCC
 10 551 GATTGCGCTG CCC..

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

1 MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
 51 DAFSGNPEGQ FFPDSYEIDA GGS DLQIYQT AYKAMQRRLN EAWESRDGL
 101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFNRLKIGM RLQTXSVIY
 15 151 GMGAAYKGKI RKADLRDTP YNTYTRGGLP PTPIALP..

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTGCGC
 51 AGCGTTTTC GCCGCGCTGC TTTTGTTC TAAGGATAAC GGCAGGGCAT
 101 ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCGTCGGT CGGCAGGAAA
 20 151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGCGGGCGGC
 201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
 251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGCGGGCAGG
 301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
 351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
 401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTAGCGGCG
 25 451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
 501 CAGTGATTTG CAGATTTACC AAACCGCCTA CAAGGCGATG CAACGCGGCC
 551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCT
 601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
 30 651 AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
 701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGTGCGGCA
 751 TACAAGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
 801 CACCTACAGC CGCGCGGTC TGCCGCCAAC CCGATTGCG CTGCCGCGCA
 851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGCGCAAAA ATACCTGTAT
 35 901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT
 951 GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTTGAAA AAATAA

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

1 MLRKLLKWSA VFLTVAADF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
 51 LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVSAW DILQMRGGR
 40 101 PDSVTQIIE GSRFSMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
 151 NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRRLEAWES RDGLPYKNP
 201 YEMLIMASLV EKETGHEADR DHVASVFNRL KIGMRLQTD PSVIYGMGAA
 251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPGSEKYL
 301 FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein encoded by yceg gene (accession P44270) of *H. influenzae*

ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

ORF7 1 MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG 55
 + G+ V+ IEG F RK ++ P + K SNE++ A ++ +
 50 yceg 102 LNSGKEVQFNVKWIIEGKTFKDWKRDLENAPHLVQTLKDKSNEEIFALLDLDPDIGQNLK 161
 ORF7 56 NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRDGLPYKNPYEMLIMAXLV 115
 N EG +PD+Y +DL++ + + M++ LN+AW R + LP NPYEMLI+A +V
 55 yceg 162 NVEGWLYPDYNTYTPKSTDLELLKRSARMKKALNKAWNERDEDLPLANPYEMLILASIV 221
 ORF7 116 EKETGHEAXXDHVASVFNRLKIGMRLQTXSVIYGMGAAYKGKIRKADLRDTPYNTYT 175
 EKETG VASVF+NRK M+LQT +VIYGMG Y G IRK DL TPYNTY
 yceg 222 EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV 281

ORF7 176 RGGLPPTPIALP 187
GLPPTPIA+P
yceg 282 IDGLPPTPIAMP 293

The complete length YCEG protein has sequence:

```

5      1 MKKFLIAILL LILILAGVAS FSYYKMTEFV KTPVNVQADE LLTIERGTTS
      51 SKLATLFEQE KLIADGKLLP YLLKLKPELN KIKAGTYSLE NVKTVQDILL
     101 LLNSGKEVQF NVKWIEGKTF KDWRKDLNA PHLVQTLKDK SNEEIFALLD
     151 LPDIGQNLEL KNVEGWLYPD TYNYPKSTD LELLKRSAR MKKALNKAWN
     201 ERDEDLPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLKAKMKLQT
     251 DPTVIYGMGE NYNGNIRKKD LETKTPYNTY VIDGLPPTPI AMPSESSLQA
     301 VANPEKTDYF YFVADGSGGH KFTRNLEHN KAVQEYLRWY RSQKNKAK

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N.*

15 *meningitidis*:

```

                                     10      20      30
orf7.pep                               MRGGRPDSVTVQIIIEGSRFESHMRKVIDATP
                                     |||
orf7a      AAYVLGVHNRHLTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIIEGSRFESHMRKVIDATP
20      70      80      90      100     110     120

                                     40      50      60      70      80      90
orf7.pep      DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLN
25      |||
orf7a      DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAMQRRLN
      130     140     150     160     170     180

                                     100     110     120     130     140     150
orf7.pep      EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXDHSVAFVFNRLKIGMRLQTXSVIY
30      |||
orf7a      EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHSVAFVFNRLKIGMRLQTDPSVIY
      190     200     210     220     230     240

                                     160     170     180
orf7.pep      GMGAAYK GKIRKADLRDTPYNTYTRGGLPPTPIALP
35      |||
orf7a      GMGAAYK GKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLYFVSKM
      250     260     270     280     290     300

40      orf7a      DGTGLSQFSHDLTEHNAVRKYILKKX
      310     320     330

```

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

```

1      ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTATCGGC
45      51 AGCCGTTTTTC GCCGCGCTGC TTTTCGTCCC TAAAGACAAC GGCAGGGCAT
     101 ACAGGATTAA AATTGCCAAA AACCAGGGTA TTTCGTGGT CCGCAGGAAA
     151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
     201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGACG TACAGACTGC
     251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCGAG
     301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
50      351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGAACACGAC ACCAAAGGCT
     401 GGAGCAATGA AAAACTGATG CGGGAAGTTG CCCCTGATG CTTACGCGGC
     451 AATCCTGAAG GGCAGTTTTT CCCCAGACAG TACGAAATCG ATGCGGGCGG
     501 CAGCGATTTA CGGATTTACC AAATCGCCTA CAAGGCGATG CAACGCCGAC
     551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
55      601 TATGAAATGC TGATTGGC GAGCCTGATC GAAAAGGAAA CAGGCGATGA
     651 AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
     701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
     751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
60      801 CACCTACACG CGCGGCGGTG TGCCGCCAAC CCCGATCGCG CTGCGCGGCA
     851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
     901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATTT
     951 GACCGAACAC AACGCCGCGG TTCGCAAATA TATTTTGAAA AAATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

```

      1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
     51 LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
    101 PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
    151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLEAWES RQDGLPYKNP
    201 YEMLIMASLI EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
    251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
    301 FVSKMDGTGL SQFSDLTEH NAAVRKYILK K*

```

A leader peptide is underlined.

10 ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

```

      10      20      30      40      50      60
or7a.pep MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
      |||
or7-1 MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
      |||
      70      80      90     100     110     120
or7a.pep HVLTAAYVLGVHNRLHTGT YRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
      |||
or7-1 HVLTAAYVLGVHNRLHTGT YRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
      |||
      130     140     150     160     170     180
or7a.pep IDATPDIEHDTKGWSNEKLM AEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
      |||
or7-1 IDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAM
      |||
      190     200     210     220     230     240
or7a.pep QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNR LKIGMRLQTD
      |||
or7-1 QRRLEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFVNR LKIGMRLQTD
      |||
      250     260     270     280     290     300
or7a.pep PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLY
      |||
or7-1 PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLY
      |||
      310     320     330
or7a.pep FVSKMDGTGLSQFSDLTEHNAAVRKYILKKX
      |||
or7-1 FVSKMDGTGLSQFSDLTEHNAAVRKYILKKX
      |||

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N. gonorrhoeae*:

```

50 orf7 MRGGRPDSVTVQIIEGSRFSHMRKV IDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQ 60
    orf7ng MRGGRPDSVTVQIIEGSRFSHMRKV IDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQ 60

55 orf7 FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMASLVEKETG 120
    orf7ng FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEKETG 120

60 orf7 HEAXXDHVASVFVNR LKIGMRLQTXSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLP 180
    orf7ng HEADRDHVASVFVNR LKIGMRLQDPSVIYGMGAAYKGKIRKADLRRDTPYNTYTGGLP 180
    orf7 PTPIALP 187

```

or_f7ng || ||||
 PTRIALPGKAAMDAAAHPSGEKYLYFVSKMDGTGLSQFSHDLTEHNAAVRKYILKK 236

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

```

5      1 MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
      51 DAFSGNPEQG FFPDSYEIDA GGSDDLQIYQT AYKAMQRRLN EAWAGRQDGL
     101 PYKNPYEMLI MASLIEKETG HEADRDHVAS VFNRLKIGM RLQTDPSVIY
     151 GMGAAYKGKI RKADLRDTP YNTYTGGGLP PTRIALPGKA AMDAAAHPSG
     201 EKYLYFVSKM DGTGLSQFSH DLTEHNAAVR KYILKK*
  
```

10 Further sequence analysis revealed a partial DNA sequence of ORF7ng <SEQ ID 37>:

```

      1 ..taccgaatca AGATTGCCAA AAATCAGGGT ATTTTCGTCGG TCGGCAGGAA
      51 ACTTGCCgaA GACCGCATCG TGTTCAGCAG GCATGTTTTG ACAGCGGCGG
     101 CCTACGTTTT GGGTGTGCAC AACAGGCTGC ATACGGGGAC gTACAGATTG
     151 CCTTCGGAAG TGTCTGCTTG GGATATCTTG CAGAAAATGC GCGGCGGCAG
    15      201 GCCGGATTCC GTTACCGTGC AGATTATCGA AGGTTTCGCGT TTTTCGCATA
     251 TGAGGAAAGT CATCGACGCA ACGCCCGACA TCGGACACGA CACCAAAGGC
     301 TGGAGCAATG AAAAAGTCTGAT GCGCGAAGTT GCGCCCGATG CCTTCAGCGG
     351 CAATCCTGAA GGGCAGTTTT TTCCCGACAG CTACGAAATC GATGCGGGCG
     401 GCAGCGATTT GCAGATTTAC CAAACCGCCT ACAAGGCGAT GCAACGCGGC
    20      451 CTGAACGAGG CATGGGCAGG CAGGCAGGAC GGGCTGCCTT ATAAAAACCC
     501 TTATGAAATG CTGATTATGG CGAGCCTGAT CGAAAAGGAA ACGGGGCATG
     551 AGGCCGACCG CGACCATGTC GCTTCCGTCT TCGTCAACCG CCTGAAAATC
     601 GGTATGCGCC TGCAAACCGA CCCGTCCGTG ATTTACGGCA TGGGTGCGGC
     651 ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGCGAC ACGCCGTACA
    25      701 aCAcctAtac gggcgggggc ttgccgccaa cccggattgc gctgccggc
     751 Aagggcgcaa tggatgccgc cgcccaccgc tccggcgaaAa aatacctgTa
     801 tttcgtgtcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT
     851 TGACCGAACA CAACGCCGCC gTcCGCAAT ATATTTTGAA AAAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 38; ORF7ng-1>:

```

30      1 ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
      51 PSEVSAWDIL QMRGGRPDS VTVQIIEGSR FSHMRKVIDA TPDIGHDTKG
     101 WSNEKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QTAYKAMQRR
     151 LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFNRLKI
     201 GMRLQTDPSV IYGMGAAYKG KIRKADLRD TPNYNTYGGG LPPTRIALPG
    35      251 KAAMDAAHP SGEKYLYFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK*
  
```

ORF7ng-1 and ORF7-1 show 98.0% identity in 298 aa overlap:

```

      10      20      30      40      50      60
orf7-1.pep KLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAE DRIVFSRHVL
40      orf7ng-1 YRIKIAKNQGISSVGRKLAE DRIVFSRHVL
      10      20      30

      70      80      90      100      110      120
orf7-1.pep TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDA
45      orf7ng-1 TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDA
      40      50      60      70      80      90

      130      140      150      160      170      180
orf7-1.pep TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR
50      orf7ng-1 TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR
      100      110      120      130      140      150

      190      200      210      220      230      240
orf7-1.pep LNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFNRLKIGMRLQTDPSV
55      orf7ng-1 LNEAWAGRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSV
      160      170      180      190      200      210

      250      260      270      280      290      300
orf7-1.pep IYGMGAAYKGKIRKADLRDTPNYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLYFVS
60
  
```

```

      |||
orff7ng-1  IYGMGAAYKKGKIRKADLRDTPYNTYTGGGLPPTRIALPGKAAMDAAAHPSGEKYLYFVS
              220      230      240      250      260      270

5      310      320      330
orff7-1.pep  KMDGTGLSQFSHDLTEHNAAVRKYILKKX
      |||
orff7ng-1    KMDGTGLSQFSHDLTEHNAAVRKYILKKX
              280      290

```

In addition, ORF7ng-1 shows significant homology with a hypothetical *E.coli* protein:

```

sp|P28306|YCEG_ECOLI_HYPOTHETICAL_38.2_KD_PROTEIN_IN_PABC-HOLB_INTERGENIC_REGION
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG_ECOLI SW: P28306 but
has 97 additional C-terminal residues [Escherichia coli] Length = 340
Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
Identities = 20/87 (22%), Positives = 40/87 (45%)

```

```

Query:      10 GISSVGRKLAEDRIVFSRHVLTAAAYVLGVHNRHTGTYRLPSEVSAWDILQKMRGGRPD 69
             G ++G +L D+I+ V + + GTYR +++ ++L+ + G+
Sbjct:      49 GRLALGEQLYADKIINRPVFWLLRIEPLDSHFAGTYRFTPMQTVREMLKLLESGKEA 108

```

```

Query:      70 SVTVQIIEGSRFSHRKVIDATPDIGH 96
             ++++EG R S K + P I H
Sbjct:     109 QFPLRLVEGMRLSDYLKQLREAPYIKH 135

```

Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
Identities = 84/155 (54%), Positives = 111/155 (71%)

```

Query:     120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRNEAWAGRDGLPYKNPYEMLIMASLIEK 179
             EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK
Sbjct:     158 EGWFWDPTWMTANTTDVALLKRAHKMKVKAVDSAWEGRADGLPYKDKNQLVTMASIIEK 217

```

```

Query:     180 ETGHEARDRHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKKGKIRKADLRDTPYNTYTGG 239
             ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT
Sbjct:     218 ETAVASERDKVASVFVNRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277

```

```

Query:     240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274
             GLPP IA PG ++ AAAHP+ YLYFV+ G
Sbjct:     276 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312

```

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 6

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>:

```

1  CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT
51 ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG
101 GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC
151 GAACGCGCAC GGCTTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC
201 GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGACAGCG GGAACGCTC
251 TGGCAACCTA TATGCTGATG TTGGAACGCA CAAATCCCC CGAAGTCGCC
301 GAACGCGCCT TGGAAATGGC CGTGTGCTG AACGCGTTG AACAGGCGGA
351 AATGATTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC
401 AAAAACGGGC GGGGTGGCTG CGGAACGTC TGAGGGAAAG AGGAAATCAG
55 451 CATCTGGACG GACGGGAAGA AGTGTGGCT CAGGCGGACG AAGGACAG

```

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

```

1  ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVERKQ QRYSEEEIKN
51 ERARLAAVGE RVNQIFLLG GETALQKGQA GTALATYMLM LERTKSPEVA
101 ERALEMAVSL NAFEQAEMII QKWRQIEPIP GKAQKRAGWL RNVLRERGNQ

```

151 HLDGREEVLA QADEGO

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

```

1  ATGTTACCTA ACCGTTTCAA AATGTTAACT GTGTTGACGG CAACCTTGAT
5  51  TGCCGGACAG GTATCTGCCG CCGGAGGCGG TGCGGGGGAT ATGAAACAGC
101 CGAAGGAAGT CGGAAAGGTT TTCAGAAAGC AGCAGCGTTA CAGCGAGGAA
151 GAAATCAAAA ACGAACGCGC ACGGCTTGCG GCAGTGGGCG AGCGGGTTAA
201 TCAGATATTT ACGTTGCTGG GAGGGGAAAC CGCCTTGCAA AAGGGGACAGG
251 CGGGAACGGC TCTGGCAACC TATATGCTGA TGTTGGAACG CACAAAATCC
301 CCCGAAGTCG CCGAACGCGC CTGGGAAATG GCCGTGTGCG TGAACGCGTT
10 351 TGAACAGGCG GAAATGATTT ATCAGAAATG GCGGCAGATT GAGCCTATAC
401 CGGGTAAGGC GCAAAACGG GCGGGGTGGC TGCGGAACGT GCTGAGGGAA
451 AGAGGAAATC AGCATCTGGA CGGACTGGAA GAAGTGCTGG CTCAGGCGGA
501 CGAAGGACAG AACCGCAGGG TGTTTTATT GTTGGCACAA GCCGCCGTGC
551 AACAGGACGG GTTGGCGCAA AAAGCATCGA AAGCGGTTTC CCGCGCGGCG
15 601 TTGAAATATG AACATCTGCC CGAAGCGGCG GTTGCCGATG TGGTGTTCAG
651 CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTTG CAGCGTTTGG
701 CGAAGCTCGA TACGGAATAA TTGCCCCCA CTTTAATGAC GTTGCGTCTG
751 ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA
801 CACCCAAAAC CTTTCGGCCG TCTGGCAGGA AATGGAAATT ATGAATCTGG
20 851 TTTCCCTGCA CAGGCTGGAT GATGCCTATG CGCGTTTGAA CGTGCTGTTG
901 GAACGCAATC CGAATGCAGA CCTGTATATT CAGGCAGCGA TATTGGCGGC
951 AAACGAAAA GAAGGTGCTT CCGTTATCGA CGGCTACGCC GAAAAGGCAT
1001 ACGGCAGGGG GACGGAGGAA CAGCGGAGCA GGGCGGCGCT AACGGCGGCG
1051 ATGATGTATG CCGACCGCAG GGATTACGCC AAAGTCAGGC AGTGGCTGAA
25 1101 AAAAGTATCC GCGCCGGAAT ACCTGTTCTG CAAAGGTGTG CTGGCGGCTG
1151 CGGCGGCTGT CGAGTTGGAC GGCGGCAGGG CGGCTTTGCG GCAGATCGGC
1201 AGGGTGCAGA AACTTCCCGA ACAGCAGGGG CGGTATTTTA CGGCAGACAA
1251 TTTGTCCAAA ATACAGATGC TCGCCCTGTC GAAGCTGCCC GATAAACGGG
1301 AGGCTTTGAG GGGGTTGGAC AAGATTATCG AAAAAACGCC TGCCGGCAGT
30 1351 AATACAGAGT TACAGGCAGA GGCATTGGTA CAGCGGTCAG TTGTTTACGA
1401 TCGGCTTGGC AAGCGGAAAA AAATGATTTT AGATCTTGAA AGGGCGTTCA
1451 GGCTTGACCC CGATAACGCT CAGATTATGA ATAATCTGGG CTACAGCCTG
1501 CTGACCGATT CCAAACGTTT GGACGAAGGT TTCGCCCTGC TTCAGACGGC
1551 ATACCAATC AACCCGGACG ATACCGCTGT CAACGACAGC ATAGGCTGGG
35 1601 CGTATTACCT GAAAGGCGAC GCGGAAAGCG CGCTGCCGTA TCTGCGGTAT
1651 TCGTTTGAAA ACGACCCCGA GCCCGAAGTT GCCGCCCAT TGGGCGAAGT
1701 GTTGTGGGCA TTGGGCGAAC GCGATCAGGC GGTGACGTA TGGACGCAGG
1751 CGGCACACCT TACGGGAGAC AAGAAAATAT GGCGGGAAC GCTCAACGCT
1801 CACGGCATCG CATTGCCCA ACCTTCCCGA AACCTCGGA AATAA

```

40 This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:

```

1  MLPNRFKMLT VLTATLIAGQ VSAAGGGAGD MKQPKEVGKV FRKQORYSEE
51  EIKNERARLA AVGERVNIQIF TLLGGETALQ KGQAGTALAT YMLMLERTKS
101 PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGKAQKR AGWLRNVLRE
45 151 RGNQHLDGLE EVLAQADEGO NRRVFLLLAQ AAVQQDGLAQ KASKAVRRAA
201 LKYEHLPEAA VADVVFVQGG REKEKAIGAL QRLAKLDEI LPPTLMTLRL
251 TARKYPEILD GFFEQTDTQN LSAVWQEMEI MNLVSLHRLD DAYARLNVLL
301 ERNPADLYI QAAILAANRK EGASVIDGYA EKAYGRGTEE QRSRAALTA
351 MMYADRRDYA KVRQWLKKVS APEYLFDKGV LAAAAVELD GGRAALRQIG
401 RVRKLPEQQG RYFTADNLSK IQMLALSKLP DKREALRGLD KIIEKPPAGS
50 451 NTELQAEALV QRSVVYDRLG KRKKMISDLE RAFRLAPDNA QIMNNLGYSL
501 LTDSKRLDEG FALLQTAYQI NPDDTAVNDS IGWAYYLKGD AESALPYLRY
551 SFENDPEPEV AAHLGEVLWA LGERDQAVDV WTQAAHLTGD KKIWRETLKR
601 HGIALPQPSR KPRK*

```

Computer analysis of this amino acid sequence gave the following results:

55 Homology with a predicted ORF from *N.meningitidis* (strain A)ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of *N.**meningitidis*:

```

60 orf9.pep      RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQORYSEEEIKNERARLA
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||
orf9a      MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEVGKVFRKQORYSEEEIKNERARLA

```

-80-

		10	20	30	40	50	
		60	70	80	90	100	110
5	orf9.pep	AVGERVNQIFTL LGGETALQKGQAGTALATYMLMLERTKSP EVAERALEMAVSLNAFEQA					
	orf9a	AVGERVNQIFTL LGXETALQKGQAGTALATYMLMLERTKSP EVAERALEMAVSLNAFEQA					
		60	70	80	90	100	110
		120	130	140	150	160	
10	orf9.pep	EMIIYQKWRQIEPIPGKAQK RAGWLRNVLRERGNQHL DGREEVLAQADEGQ					
	orf9a	EMIIYQKWRQIEPIPGKAQK RAGWLRNVLRERGNQHL DGLEEXLAQADEXQNR RVFLLAQ					
		120	130	140	150	160	170
15	orf9a	AAVQQDGLAQKASKAVRR AALRYEHLPEAAVADV VFSVQXREKEKAIGALQRLAKLDTEI					
		180	190	200	210	220	230

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

1	ATGTTACCG	CCCCTTTAC	CATTTTATCT	GTGCTCGCG	CAGCCCTGCT
51	TGCCGGGCAG	CGGTATGCCG	CCGGCGCGGC	GGATGCGAAG	CCGCCGAAGG
101	AAGTCGGA	GGTTTTAGA	AAGCAGCAGC	GTTACAGCGA	GGAAGAAATC
151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAGCGGG	TTAATCAGAT
201	ATTTACGTTG	CTGGGANGGG	AAACCGCCTT	GCAAAAGGGG	CAGGCGGGAA
251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCGAA
301	GTCGCCGAAC	GCGCCTTGA	AATGGCCGTG	TCNCTGAACG	CGTTTGAACA
351	GGCGGAAATG	ATTTATCAGA	AATGGCGGCA	GATTGAGCCT	ATACCGGGTA
401	AGGCGCAAAA	ACGGGCGGGG	TGGCTGCGGA	ACGTGCTGAG	GGAAAGAGGA
451	AATCAGCATC	TAGACGGACT	GGAAGAANTG	CTGGCTCAGG	CGGACGAANG
501	ACAGAACCGC	AGGGTGT TTT	TATTGTTGGC	ACAAGCCGCC	GTGCAACAGG
551	ACGGGTTGGC	GCAAAAAGCA	TCGAAAGCGG	TTCCGCCGCG	GGCGTTGAGA
601	TATGAACATC	TGCCCGAAGC	GGCGGTTGCC	GATGTGGTGT	TCAGCGTACA
651	GGNACGCGAA	AAGGAAAAGG	CAATCGGAGC	TTTGACGCGT	TTGGCGAAGC
701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
751	CGCAAAATATC	CCGAAATACT	CGACGGCTTT	TTGAGCAGAG	CAGACACCCA
801	AAACCTTTCC	GCGCTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
851	TGCACAGGCT	GCTGATGCCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACCG
901	AATCCGAATG	CAGACCTGTA	TATTGAGGCA	GCGATATTGG	CGGCAACCCG
951	AAAAGAANGT	GCTTCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
1001	GGGGGACGGG	GGAACAGCGG	GGCAGGGCGG	CAATGACGGC	GGCGATGATA
1051	TATGCCGACC	GAAGGATTA	CACCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	TGTGCTGGCG	GCTGCGGCGG
1151	CTGTTCAGTT	GGACNCGCGC	AGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
1201	CGGAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
1251	CAAAATACAG	ATGTTGCCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAGGCTT
1301	TGAGGGGGTT	GGACAAGATT	ATCGAAAAAC	CGCCTGCCGG	CAGTAATACA
1351	GAGTTACAGG	CAGAGGCATT	GGTACAGCGG	TCAGTTGTTT	ACGATCGGCT
1401	TGGCAAGCGG	AAAAAATGA	TTTCAGATCT	TGAAAGGGCG	TTCAGGCTTG
1451	CACCCGATAA	CGCTCAGATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
1501	GATTCCAAC	GTTTGGACGA	AGGCTTCGCC	CTGCTTCAGA	CGGCATACCA
1551	AATCAACCCG	GACGATACCG	CTGTCAACGA	CAGCATAGGC	TGGGCGTATT
1601	ACCTGAAANG	CGACGCGGAA	AGCGCGCTGC	CGTATCTGCG	GTATTCTGTT
1651	GAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTGT
1701	GGCATTGGGC	GAACGCGATC	AGGCGGTTGA	CGTATGGACG	CAGGCGGCAC
1751	ACCTTACGGG	AGACAAGAAA	ATATGGCGGG	AAACGCTCAA	ACGTCACGGC
1801	ATCGCATTGC	CCCAACCTTC	CCGAAAACCT	CGGAAATAA	

55 This encodes a protein having amino acid sequence <SEQ ID 44>:

1	MLPARFTILS	VLAAALLAGO	AYAAGAADAK	PPKEVGKVER	KQORYSEEEI
51	KNERARLAAV	GERVNQIFTL	LGXETALQKG	QAGTALATYM	LMMLERTKSP
101	VAERALEMAV	SLNAFEQAEM	IYQKWRQIEP	IPGKAQKRAG	WLRNVLRERG
151	NQHL DGLEEX	LAQADEXQNR	RVFLLLAQAA	VQDGLAQKA	SKAVRRRAALR
201	YEHLPEAAVA	DVVFVSQXRE	KEKAIGALQR	LAKLDTEILP	PTLMTLRRLTA
251	RKYPEILDGF	FEQTDTONLS	AVWQEMEIMN	LVSLHRLDDA	YARLNVLLER
301	NPADLYIQA	AILAANRKEK	ASVIDGYAEK	AYGRGTGEQR	GRAAMTAAMI
351	YADRRDYTKV	RQWLKKVSAP	EYLFDKGVLA	AAAAVELDXG	RAALRQIGRV
401	RKLPEQQGRY	FTADNLSKIQ	MFALSKLPDK	REALRGLDKI	IEKPPAGSNT
451	ELQAEALVQR	SVVYDRLGKR	KKMISDLERA	FRLAPDNAQI	MNNLGYSLLS
501	DSKRLDEGFA	LLQTAYQINP	DDTAVNDSIG	WAYYLLKXDAE	SALPYLRYSF
551	ENDPEPEVAA	HLGEVLWALG	ERDQAVDVWT	QAAHLTGDKK	IWRETLKRHG

601 IALPQPSRKPK RK*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVFRKQORYSEEEIKNERARLA
	orf9-1	10 20 30 40 50 60	MLPNRFKMLTTLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFRKQORYSEEEIKNERARLA
10	orf9a.pep	60 70 80 90 100 110	AVGERVNQIFTLGKETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
	orf9-1	70 80 90 100 110 120	AVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
15	orf9a.pep	120 130 140 150 160 170	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLGLEEXLAQADEXQNRVFLLLAQ
	orf9-1	130 140 150 160 170 180	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLGLEEVLAQADEGQNRVFLLLAQ
20	orf9a.pep	180 190 200 210 220 230	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTEI
25	orf9-1	190 200 210 220 230 240	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVFSVQGREKEKAIGALQRLAKLDTEI
30	orf9a.pep	240 250 260 270 280 290	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
	orf9-1	250 260 270 280 290 300	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
35	orf9a.pep	300 310 320 330 340 350	ERNPNADLYIQAAILAANKRKEASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYT
	orf9-1	310 320 330 340 350 360	ERNPNADLYIQAAILAANKRKEGASVIDGYAEKAYGRGTTEEQRSAALTAAMMYADRRDYA
40	orf9a.pep	360 370 380 390 400 410	KVRQWLKKVSAPEYLFDKGVLA AAAA VELDXGRAALRQIGRVKRLPEQQGRYFTADNLSK
	orf9-1	370 380 390 400 410 420	KVRQWLKKVSAPEYLFDKGVLA AAAA VELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK
45	orf9a.pep	420 430 440 450 460 470	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQSVVYDRLGKRKKMISDLE
	orf9-1	430 440 450 460 470 480	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQSVVYDRLGKRKKMISDLE
50	orf9a.pep	480 490 500 510 520 530	RAFRLAPDNAQIMNNLGYSLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
55	orf9-1	490 500 510 520 530 540	RAFRLAPDNAQIMNNLGYSLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
60	orf9a.pep	540 550 560 570 580 590	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
	orf9-1	550 560 570 580 590 600	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
65	orf9a.pep	600 610	HGIALPQPSRKPRKX
	orf9-1	610	HGIALPQPSRKPRKX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N.gonorrhoeae*:

5	Orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEGVKVFRRKQORYSEEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVLPKEVGKVLRRKHRRYSEEEIKNERAR	58
	orf9	LAAVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
10	orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
	orf9	QAEMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPAQSDYVHQPMIFLLL	178

15 The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

	1	MIMLPARFTI	LSVLAAALLA	GOAYAAGAAD	VELPKVEGVK	LRKHRRYSEE
	51	EIKNERARLA	AVGERVNRVF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
20	101	PEVAERALEM	AVSLNAFEQA	EMIQKWRQI	EPIPGEAQKP	AGWLRNVLKE
	151	GGNPHLDRLE	EVPAQSDYVH	QPMIFLLLVO	AAVQHGGVAQ	KPSKAVRPAA
	201	YNYEVLPEA	GADAVFCVQG	PQYEKAIQSF	PPCGRNPQTE	NIAPPFNELF
	251	RPTARPISPK	LLQRFFRTEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

	1	ATGTTACCCG	CCCGTTTTCAC	TATTTTATCT	GTCCTCGCAG	CAGCCCTGCT
	51	TGCCGGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGA	GGTTTTAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
30	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
	201	GTTTACGCTG	TTGGGCGGTG	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	GGCGGAAATG	ATTTATCAGA	AATGgcggca	gacgcagcct	ataCcggggtg
	401	agggcgcaaaa	accgGcgggg	tggctgcgga	acgtattgaa	ggaagggGGA
35	451	aaTCAGCATC	TGGAcggggtt	gaaagaggTG	CtggcgcaAT	cggacgatGT
	501	GCAAAAacgc	aggaTATTTT	TGCTGCTGGT	GCAAGCCGCC	GTGCagcagg
	551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTCGcgtg	GGcgttgaAG
	601	TATGAACATC	TGCCcgaagc	ggcggTTGCC	GATGcggTGT	TCGGCGTACA
	651	GGGACGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGCAGCGT	TTGGCGAAGC
40	701	TCGATACGGA	AATATTGCC	CCCACTTTAA	TGACGTTCG	TCTGACTGCA
	751	CGCAATATC	CCGAAATACT	CGACGGCTTT	TTCGAGCAGA	CAGACACCCA
	801	AAACCTTTTC	GCCGCTCTGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
	851	TGCGTAAGCC	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACAC
	901	AACCCGAATG	CAACCTGTA	TATTCAGGCG	GCGATATTGG	CGGCAAACCG
45	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCagggcgg	cAATgacggc	GGCGATGATA
	1051	TATGCCGACC	GCAGGGATTA	CGCCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	CGTCTGGCG	GCTGCGGCGG
	1151	CTGCCGAATT	GGACGGAGGC	CGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
50	1201	CGGAACTTC	CGAACAAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTGTTC
	1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAAGCCC
	1301	TGATCGGGCT	GAACAACATC	ATCGCCAAAC	TTTCGGCGGC	GGGAAGCACG
	1351	GAACCTTTGG	CGGAAGCATT	GGCACAGCGT	TCCATTATTT	ACGaacAGTT
	1401	cggCAAACGG	GGAAAAATGA	TTGCCGACCT	tgaAACcgcg	CTCAAACTTA
55	1451	CGCCCGATAA	TGCACAAATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
	1501	GATTCCAAAC	GTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CCGTTAACGA	CAGCATAGGC	TGGGCGTATT
	1601	ACCTGAAAGG	CGACcggaA	AGCGCGCTGC	CGTATCTGcg	gtattcgttt
	1651	gAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTGTG

5	1	MLPARFTILS	VLAALLAGO	AYAAGAADVE	LPKEVGKVL	KHRRYSEEEI
	51	KNERARLAAV	GERVNRVFTL	LGGETALQKG	QAGTALATYM	LMLERTKSPE
	101	VAERALEMAV	SLNAFEQAEM	IYQKWRQIEP	IPGEAAQKPA	WLRNVLKEGG
	151	NQHLDGLKEV	LAQSDDVQKR	RIFLLLVQAA	VQQGGVQAQK	SKAVRRAALK
10	201	YEHLPEAAVA	DAVGVQGRE	KEKATEALQR	LAKLDTEILP	PTLMTLRLTA
	251	RKYPEILDGF	FEQTDTONLS	AVWQEMEIMN	LVSLRKPDPA	YARLNVLLEH
	301	NPNANLYTQA	AILAANRKEG	ASVIDGYAEK	AYGRGTGEQR	GRAAMTAAMI
	351	YADRRDYAKV	RQWLKKVSAP	EYLFDKGVLA	AAAAAELDGG	RAALRQIGRV
15	401	RKLPEQQGRY	FTADNLSKIQ	MLALSCLPKD	REALIGLNNI	IAKLSAAGST
	451	EPLAEALAQR	SIIYEQFGKR	GKMIADLETA	LKLTDPDQAI	MNNLGYSLLS
	501	DSKRLDEGFA	LLQTAYQINP	DDTAVNDSIG	WAYYLLKGD	SALPYLRYSF
	551	ENDPEPEVA	HLGEVLWALG	ERDQAVDVWT	QAAHLRGDKK	IWRETLKRYG
	601	IALPEPSRKP	RK*			

		10	20	30	40	50	60
20	orf9-1.pep	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEGVKFRKQORYSEEEIKNERARLA					
	orf9ng-1	MLPARFTILSVLAAALLAGQAYAAG--AADVELPKEGVKVLKRRRYSEEEIKNERARLA					
		10	20	30	40	50	
25	orf9-1.pep	AVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMVSLNAFEQA	70	80	90	100	110
	orf9ng-1	AVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMVSLNAFEQA					
		60	70	80	90	100	110
30	orf9-1.pep	EMIQKWRQIEPIPGKAQKRAEWLRNVLRRGNQHLGLLEEVLAQADEQNRRVFLLLAQ	130	140	150	160	170
	orf9ng-1	EMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNQHLDGLKEVLAQSDDVQKRIFLLLVQ					
35		120	130	140	150	160	170
40	orf9-1.pep	AAVQQDGLAQKASKAVRRAALKYEHLPAAVADVVFVSGREKEKAIGALQRLAKLDTEI	190	200	210	220	230
	orf9ng-1	AAVQQGGVAQKASKAVRRAALKYEHLPAAVADAVFGVQGREKEKAIEALQRLAKLDTEI					
		180	190	200	210	220	230
45	orf9-1.pep	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL	250	260	270	280	290
	orf9ng-1	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKPDDAYARLNVLL					
		240	250	260	270	280	290
50	orf9-1.pep	ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQRSRAALTAAMMYADRRDYA	310	320	330	340	350
	orf9ng-1	EHNPNANLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQGRAAMTAAMIYADRRDYA					
		300	310	320	330	340	350
55	orf9-1.pep	KVRQWLKKVSAPEYLFDKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK	370	380	390	400	410
	orf9ng-1	KVRQWLKKVSAPEYLFDKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK					
		360	370	380	390	400	410
60	orf9-1.pep	IQMLALSCLPKREALRGLDKIIEKPPAGSNTLQAEALVQSVVYDRLGKRRKMISDLE	430	440	450	460	470
	orf9ng-1	IQMLALSCLPKREALIGLNNYIAKLSAAGSTEPLAEALAQRSIIEYQFGKRGKMIADLE					
65		420	430	440	450	460	470
		480	490	500	510	520	530
							540

10

15

20

25

30

35

40

45

50

55

60

65

70

BNSDOCID: <WO__9924578A2_1_>

-85-

Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390

Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513
 +I+Y+ G L A++L P+N N LGYSLL +R++E L++

5 Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYLGYSLLLWYGKERVEEAEELIK 450

Query: 514 TAYQINPDDTAVNDSIGWAYYYLKGD AESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572
 A + +P++ A DS+GW YYLKGD E A+ YL + E +P V H+G+VL +G +

10 Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKGDYERAMQYLLKALREAYDDPVVNEHVGDVLLKMGYK 510

Query: 573 DQAVDVWVWQAHLRGDKK 590
 ++A + + +A L + K

Sbjct: 511 EEARNYYERALKLLEEGK 528

- 15 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 7

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

20 1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
 51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
 101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
 151 TGGGCGATTA TCGTTTAAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
 201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
 25 251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
 301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CG₂CTGGGCG
 351 GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
 401 GCATTGTTCG CCTCCGTAGA ATTGCGCCAG GCACCTGGC TGGGTGGAT
 451 TACCGACCTC AGCCGCGCCG ACCCTACTA CATCTGCC ATCATTATGG
 501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCGCGCCG GACCGACCG
 30 551 ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTCT CsGwCrTGTT
 601 CTTCTTCTTC CCTGCCGgks TGGTATTGTA CTGGGTAGTC AACAACTCC
 651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
 701 GCCCAAGGCG AAGTCGTTTC CTAA

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

35 1 .NLYAGPQTTTS VIANIADNLQ LAKDYGKVHW FASPLFWLLN QLHNIIGNWG
 51 WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQQ
 101 QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLWGI
 151 TDLRADPHY ILPIIMAATM FAQTYLNPPP TDPMAKMMK IMPLVFSXXF
 201 FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS *

- 40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

1 ATGGATTTTA AAAGACTCAC GCGGTTTTTC GCCATCGCGC TGGTGATTAT
 51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
 101 AACAGGCAGC ACAACAACAG GCCGTAACCG CTTCCGCCGA AGCCGCGCTC
 151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTGAT
 45 201 TGATGAAAAA AGCGCGGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
 251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGCGCA CGGCAAAGAA
 301 TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
 351 TCTAAAAGGC ATCGGCTTTA GCGCACCAGG AAAACAGTAC AGCTTGGAAG
 401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CGGTCTGAAA
 50 451 ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCT
 501 CTTGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
 551 ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG TTACTTTACC
 601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
 651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
 55 701 CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC
 751 CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAC AAAGCGTTTG
 801 CGCCGCGAGG GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGTGT
 851 ACGACCAAG CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA
 901 GCCGAAGCCT CCATCAACCT CTACGCCGGC CCGCAGACCA CATCCGTGAT
 60 951 CGCAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAGTAC

-86-

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160

1001 ACTGGTTCGC CTCCTGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAATGC
1151 GTGCCGCCGC ACCCAAATG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACAAACAGG GATGATGCAG CTTTACACAG ACGAGAAAAT
1251 CAACCCGCTG GCGGGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401 CCCCATCATT ATGGCGGCAA CGATGTTTCG CCAAACCTTAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCGTGGTT
1501 TTCTCCGTCA TGTTCTTCTT CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCAGCAGCA
1601 TCGAAAAACA ACGCGCCAA GCGGAAGTCG TTTCTCTAA

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

15
20
25
30
35
40
45
50

1 MDFKRLTAFF AIALVIMIGW EKMFPPTPKPV PAPOQAAQQQ AVTASAEAL
51 APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFILFGDGKE
101 YTYVAQSELL DAQGNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKVYTFTEG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVYTT PEGNFQKVSF SLDLDDAKSG KSEAHEYIRKT PTGWLGMIEH
251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
301 AEASINLYAG PQTTSVIANI ADNLQLAKDY GKVHWFASPL FWLLNQLHNI
351 IGWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEYGDD
401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451 WLGWITDLR ADPYYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
501 FSVMFFFFPA GLVLYVWVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

ORF11 2 LYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61
LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K
60K 324 LYAGPKIQSKLKLSPGLELTVDYGFLWFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIK 383
ORF11 62 AVLYPLTNASYRSMAMRAAAPKLQAIKEYGDDRRXXXXXXXXXXLYTDEKINPLGGCLPM 121
+ +PL+ ASYRSMA+MRA APKL A+KE++GDDR LY EKNINPLGGCLP+
60K 384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443
ORF11 122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLRADPYYILPIIIMATMFAQTYLNPPPT 181
L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P
60K 444 LVQMPVFLALYVWVLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQRLNPTTP 503
ORF11 182 DPMQAKMMKIMPLVXXXXXXXXXPAGXVLYVWVNNLLTIAQQWHINRSIE 230
DPMQAK+MK+MP++ PAG VLYVWVNN L+I+QQW+I R IE
60K 504 DPMQAKVMKMPPIIFTFFFLWFPAGLVLYVWVNNCLSSISQQWYITRRIE 552

45 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N.*

meningitidis:

50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

orf11.pep NLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKE
orf11a IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQLKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKE
280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510 520 530 540 550 560 570 580 590 600 610 620 630 640 650 660 670 680 690 700 710 720 730 740 750 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900 910 920 930 940 950 960 970 980 990 1000

-87-

		100	110	120	130	140	150
orfl1.pep		KYGD	DRMA	QQQAM	QLYT	DEKIN	PLGG
orfl1a		KYGD	DRMA	QQQAM	QLYT	DEKIN	PLGG
5		400	410	420	430	440	450
orfl1.pep		TDL	SRAD	PYYIL	PIIMA	ATMFA	QTYLN
10		TDL	SRAD	PYYIL	PIIMA	ATMFA	QTYLN
		460	470	480	490	500	510
orfl1.pep		WVV	NNLL	TIAQ	QWHIN	RSIEK	QRAQ
15		WVV	NNLL	TIAQ	QWHIN	RSIEK	QRAQ
		520	530	540			

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

	1	ANGG	ATTTTA	AAAG	ACTCAC	NGNG	TTTTTC	GCCAT	CGCAC	TGGT	GATTAT
20	51	GATC	GGATNG	NAA	ANGATGT	TCCC	ACTCC	GAAG	CCCGTC	CCCG	CGCCCC
	101	AAC	GACGGC	ACA	ACAACAG	GCCG	TAANCG	CTTC	CGCCGA	AGCC	CGCGCTC
	151	GCG	CCCGNAN	CGCC	GATTAC	CGTA	ACGACC	GAC	ACGGTTC	AAG	CCGTCAT
	201	TGAT	GAAAAA	AGCG	GCGACC	TGCG	CCCGCT	GACC	CTGCTC	AAAT	ACAAAG
25	251	CAAC	CGGCGA	CNAA	AATAAA	CCGT	TATCC	TGTT	TGGCGA	CGG	CAANAA
	301	TAC	ACCTACN	TCGC	CCANTC	CGAA	CTTTG	GAC	GCGCAGG	GCA	ACAACAT
	351	TCT	AAAAGGC	ATCG	GCTTTA	GCG	CACGAA	AAA	ACAGTAC	AGCT	TGGAAG
	401	GCG	ACAAAGT	TGA	AGTCCGC	CTGA	GCGCAC	CTGA	AACACG	CGGT	CTGAAA
	451	ATC	GACAAAG	TTTA	TACTTT	CAC	CAAAGGC	AGCT	TATCTGG	TCA	ACGTCGG
30	501	CTT	CGACATC	GCCA	ACGGCA	GCGG	TCAAAC	CGCA	ACCTG	AGCG	CGGACT
	551	ACCG	CATCGT	CCGC	GACCAC	AGCG	AACCGG	AGGT	TCAAGG	CTAC	TTTACC
	601	CAC	TCTTACG	TCGG	CCCTGT	TGTT	TATACC	CCT	GAAAGGCA	ACT	TCCAAAA
	651	AGT	CAGCTTC	TCCG	ACTTGG	ACG	ACGATGC	CAAN	TCCGN	AAAT	CCGAGG
	701	CCG	AATACAT	CCG	CAAAACC	CN	GACGGCT	GGCT	CGGCAT	GATT	GAACAC
35	751	CAC	TTTATGT	CCAC	TGGAT	CCT	CCAACCC	AAAG	GCGGAC	AAAG	CGTTTG
	801	CGC	CGCTGGC	GACT	GCNGTA	TNG	ACATCAA	ACG	CCGCAAC	GACA	AGCTGT
	851	AC	AGCACCAG	CGT	CAGCGTG	CCT	TAGCCG	CTAT	TCAAAA	CGGT	GCGAAA
	901	TC	NAAGCCT	CCAT	CAACCT	CTAC	GCCGGC	CCAC	AGACCA	CATC	NGTTAT
	951	CG	CAAAACATC	GCC	GACAACC	TGCA	ACTGGN	CAA	AGACTAC	GGCA	AAAGTAC
40	1001	ACT	GTTTCGC	CTCC	CCCCCTC	TTTT	GGCTTT	TGA	ACCAACT	GCAC	AACATC
	1051	ATC	GGCAACT	GGGG	CTGGG	GATT	ATCGTT	TTA	ACCATCA	TCGT	CAAAGC
	1101	CGT	ACTGTAT	CCAT	TGACCA	ACGC	CTCTTA	CCGT	TCGATG	GCG	AAAAATGC
	1151	ITG	CGCCCGC	GCC	CAAACTG	CAAG	CCATCA	AAGA	GAAATA	CGGC	AGCGAC
	1201	CGT	ATGGCGC	AGCA	ACAAGC	CAT	GATGCAG	CTTT	ACACAG	ACG	AGAAAAT
	1251	CA	ACCGCTG	GGCG	GCTGCC	TGCC	TATGCT	GTT	GCAAAATC	CCCG	TCTTCA
45	1301	TCG	GATTGTA	TTGG	GATTG	TTGC	CTCCG	TAGA	ATTGCG	CCAG	GCACCT
	1351	TGG	CTGGGT	GGAT	TACCGA	CCT	CAGCCG	GCC	GACCCNT	ACT	ACATCCT
	1401	GCCC	ATCATT	ATGG	CGGCAA	CGAT	GTTTCGC	CCAA	ACCTAT	CTGA	ACCCGC
	1451	CGCC	GACCGA	CCG	ATGCAG	GCG	AAAATGA	TGAA	AATCAT	GCCT	TTGGTT
50	1501	NTNT	CNNNNA	NGTT	CCTCNN	CTTC	CCTGCC	GGT	CTGGTAT	TGT	ACTGGGT
	1551	GAT	CAACAAC	CTCT	GACCA	TCGC	CCAGCA	ATGG	CACATC	AACC	CGCAGCA
	1601	TCG	AAAAACA	ACGC	GCCCCAA	GGCG	AAGTCG	TTTC	TCTAA		

This encodes a protein having amino acid sequence <SEQ ID 54>:

	1	XDFK	RLTXFF	AIAL	VIMIGX	XXMF	PTPKPV	PAPO	QTAQOO	AVX	ASAEAL
55	51	APXX	PITVTT	DTVQ	AVIDEK	SGDL	RRLTLL	KYKAT	GDXNK	PFIL	FGDGKX
	101	YTYX	AXSELL	DAQ	NNILKG	IGFS	APKKQY	SLEG	DKVEVR	LSAP	PETRGLK
	151	IDKV	YFTKG	SYLV	NVRFDI	ANGS	GQTANL	SADY	RIVRDH	SEPE	GQGYFT
	201	HSYV	GPVYVT	PEGN	FQKVSF	SDL	DDAXSG	KSEA	EYIRKT	XTGW	LGMIEH
	251	HFM	STWILQP	KGGQ	SVCAAG	DCXX	DIKRNR	DKLY	STSVSV	PLAA	IQNGAK
60	301	SXAS	INLYAG	PQTT	SVIANI	ADNL	QLXKDY	GKVH	WFASPL	FWLL	NQLHNI
	351	IGNW	GWAIIV	LTII	VKAVLY	PLTN	ASYRSM	AKMR	AAAPKL	QAIKE	KYGDD
	401	RMA	QQQAMMQ	LYT	DEKINPL	GGCL	PMLLQI	PVFI	GLYWAL	FAS	VELRQAP
	451	WLGW	ITDLR	ADPY	YILPII	MAAT	MFAQTY	LNPP	PTDPMQ	AKMM	KIMPLV
	501	XSXX	FFXFPA	GLVLY	WVINN	LLTIA	QQWHI	NRSIE	KQRAQ	GEVVS*	

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

65	10	20	30	40	50	60
----	----	----	----	----	----	----

5	orf11a.pep	XDFKRLTXFFAIALVIMIGXXXMFPTPKVPVPAQQTAAQQQAVXASAEAAALAPXXPITVTT
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVPAQQAQQQAVTASAEAAALAPATPITVTT 10 20 30 40 50 60
10	orf11a.pep	DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFILFGDGKXYTYXAXSELLDAQGNNILKG
	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFILFGDGKEYTYVAQSELLDAQGNNILKG 70 80 90 100 110 120
15	orf11a.pep	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL 130 140 150 160 170 180
20	orf11a.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVYTPEGNFQKVSFSDLDDDAKSGKSEAIEYIRKT
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVYTPEGNFQKVSFSDLDDDAKSGKSEAIEYIRKT 190 200 210 220 230 240
25	orf11a.pep	XTGWLGMIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGDCNIDIKRRNDKLYSTSVSVPLAAIQNGAK 250 260 270 280 290 300
30	orf11a.pep	SXASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV :
	orf11-1	AEASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV 310 320 330 340 350 360
35	orf11a.pep	LTIIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKEKYGDDRMAMQQAAMQLYTDEKINPL
	orf11-1	LTIIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKEKYGDDRMAMQQAAMQLYTDEKINPL 370 380 390 400 410 420
40	orf11a.pep	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIIMAAATMFAQTY
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIIMAAATMFAQTY 430 440 450 460 470 480
45	orf11a.pep	LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ 490 500 510 520 530 540
50	orf11a.pep	GEVVSX
	orf11-1	GEVVSX

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N. gonorrhoeae*:

65	Orf11	NLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT 	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT 	60

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

	1	MDFKRLTAFF	AIALVIMIGW	EKMFFTPKPV	PAPOQAAQKO	AATASAEAL
	51	APATPITVTT	DTQVAVIDEK	SGDLRLTL	KYKATGDENK	PFVLFQDGKE
60	101	YTYVAQSELL	DAQGNILKG	IGFSAPKKQY	TLNGDTVEVR	LSAPETNGLK
	151	IDKVYTFTKD	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDR	SEPEGQGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDAKSG	KSEAEYIRKT	PTGWLGMIEH
	251	HFMSTWILQP	KGGQNVCAQG	DCRIDIKRRN	DKLYASVSV	PLTAIPTRPG
	301	KPKMAVNLYA	GPOTTSVIAN	IADNLQAKD	YGVXHWFSAP	LFWLLNQLHN

351 IIGNWGWAI VLTIIIVKAVL YPLTNASYRS MAKMRAAAPK LQTIKEYGD
 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLO IPVFIGLYWA LFASVELRQA
 451 PWLGWITDLS RADPYYILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL
 501 VFSVMFFFFP AGLVLYWVVN NLLTIAQQWH INRSIEKQRA QGEVVS*

5 ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

		10	20	30	40	50	60
	orf11ng-1.pep	MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPA	QQAQQAATASAEAA	LAPATPITVTT			
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPA	QQAQQAATASAEAA	LAPATPITVTT			
10		10	20	30	40	50	60
	orf11ng-1.pep	DTVQAVIDEKSGDLRRLTLLKYKATG	DNKPFVLF	FGDGKEYTYVAQ	SELLDAQ	GNNILKG	
	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATG	DNKPFVLF	FGDGKEYTYVAQ	SELLDAQ	GNNILKG	
15		70	80	90	100	110	120
	orf11ng-1.pep	DTVQAVIDEKSGDLRRLTLLKYKATG	DNKPFVLF	FGDGKEYTYVAQ	SELLDAQ	GNNILKG	
	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATG	DNKPFVLF	FGDGKEYTYVAQ	SELLDAQ	GNNILKG	
20		130	140	150	160	170	180
	orf11ng-1.pep	IGFSAPKKQYTLNGD	TVVRLSAPETNGLKIDKVY	TFTKDSYLVN	VRFDIANGSGQTANL		
	orf11-1	IGFSAPKKQYTLNGD	TVVRLSAPETNGLKIDKVY	TFTKDSYLVN	VRFDIANGSGQTANL		
25		190	200	210	220	230	240
	orf11ng-1.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVY	TPEGNFQKVSFSD	LDDDAKSGKSEA	EYIRKT		
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVY	TPEGNFQKVSFSD	LDDDAKSGKSEA	EYIRKT		
30		250	260	270	280	290	300
	orf11ng-1.pep	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGD	CRIDIKRRNDKLYSASVS	PLTAIPTRGP			
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGE	CNIDIKRRNDKLYSTSVS	PLAAIQN-GA			
35		310	320	330	340	350	360
	orf11ng-1.pep	KPKMAVNLYAGPQTTSVIANIADNLQ	LAKDYGKVVHWFASPLFWLLNQLHNI	IIGNWGWAI			
	orf11-1	KAEASINLYAGPQTTSVIANIADNLQ	LAKDYGKVVHWFASPLFWLLNQLHNI	IIGNWGWAI			
40		370	380	390	400	410	420
	orf11ng-1.pep	VLTIIIVKAVLYPLTNASYRSMAKMRAAPK	LQTIKEYGDD	MAQQQAMMQLYKDEKINP			
	orf11-1	VLTIIIVKAVLYPLTNASYRSMAKMRAAPK	LQAIKEYGDD	MAQQQAMMQLYTDEKINP			
45		430	440	450	460	470	480
	orf11ng-1.pep	LGGCLPMLLOIPVFIGLYWALFASVELRQA	PWLGWITDLSRADPYYILPI	IMAATMFAQT			
	orf11-1	LGGCLPMLLOIPVFIGLYWALFASVELRQA	PWLGWITDLSRADPYYILPI	IMAATMFAQT			
50		490	500	510	520	530	540
	orf11ng-1.pep	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFP	AGLVLYWVVNNLLTIAQQWHINRSIEKQRA				
	orf11-1	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFP	AGLVLYWVVNNLLTIAQQWHINRSIEKQRA				
55		540	550	560	570	580	590
	orf11ng-1.pep	QGEVVSX					
	orf11-1	QGEVVSX					
60		540					

65 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

-91-

ID 60IM PSEPU STANDARD; PRT; 560 AA.
AC P25754;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE 60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES Init1: 1074 Initn: 1293 Opt: 1103
Smith-Waterman score: 1406; 41.5% identity in 574 aa overlap

```

10      10      20      30      40
orfl1ng-1.pep MDFKR---LTAFFAIALVIMIGW-----EKMFP-----PKPVPAQQAAQKQ
p25754      MDIKRTILIAALAVVSYVMVLKWNDDYGQAALPTQNTAASTVAPGLPDGVPAGNNGASAD
15      10      20      30      40      50      60
orfl1ng-1.pep      50      60      70      80      90
p25754      AATASAEAALAPATPIT-----VTTDTVQAVIDEKSGDLRRLTLLKYKATGDE-NKPF
20      70      80      90      100      110      120
p25754      VPSANAESSPAELAPVALSKDLIRVKTDLVLELAIDPVGGDIVQLNLPKYPRRQDHPNIPF
25      100      110      120      130      140
orfl1ng-1.pep      VLFGDGKEYTYVAQSELLDAQNNILKGIG---FSAPKKQYTL-NGD---TVEVRLSAPE
p25754      QLFDNNGGERVYLAQSGLTGTGDPDA-RASGRPLYAAEQSYQLADGQEQLVVDLKFSS---
30      130      140      150      160      170
orfl1ng-1.pep      150      160      170      180      190      200
p25754      TNLKIDKVYTFTKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY
35      180      190      200      210      220      230
p25754      DNGVNYIKRFSEFKRGEYDLNWSYLIIDNQSGQAWNGNMFAQLKRDSGDPSSSTATGTATY
40      210      220      230      240      250      260
orfl1ng-1.pep      VGPVYVTPEGNFQKVSFSDLDDDAKSGKSEAERYIKRTPGTWGLMIEHHFMSTWILQPKGG
p25754      LGAALWTASEPYKKVSMKDID---KGSLE-----NVSGGWVAWLQHYFVTAWI-PAKSD
45      240      250      260      270      280
orfl1ng-1.pep      270      280      290      300      310      320
p25754      QNVCAQGDCRIDIKRRNDKLYSASVSVPLTAIPTRGPKPKMAVNLYAGPQTTSVIANIAD
50      290      300      310      320      330
p25754      NNV-----VQTRKDSQGYIIGYTGPVISVPA-GGKVETSALLYAGPKIQSKLKELSP
55      330      340      350      360      370      380
orfl1ng-1.pep      NLQLAKDYGKVHWF-ASPLFWLLNQLHNIIGNWGWAVVLTIIIVKAVLYPLTNASYSMA
p25754      GLELTVDYGFL-WFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIKGLFFPLSAASYSMA
60      340      350      360      370      380      390
orfl1ng-1.pep      390      400      410      420      430      440
p25754      KMRAAAPKLQTIKEKYGDDRMAQQQAMMQLYKDEKINPLGGCLPMLLQIPVFIGLYWALF
65      400      410      420      430      440      450
p25754      RMRVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPILVQMPVFLALYWVLL
70      450      460      470      480      490      500
orfl1ng-1.pep      ASVELRQAPWLGWITDLRADPYIILPIIIMAAATMFAQTYLNPPPTDFMQAKMMKIMPLVF
p25754      ESVEMRQAPWILWITDLSEIKDPFFILPIIMGATMFIQRLNPTPPDFMQAKVMKMMPIIF
75      460      470      480      490      500      510
orfl1ng-1.pep      510      520      530      540
p25754      SVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGEVVSX
80      520      530      540      550      560
p25754      TFFFLWFPAGLVLYWVVNNCLSSISQWYITRRIEAAATKAAA
```

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```

1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTNG
151 TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251 ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

15 This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```

1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVXY RGTXWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```

20 1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTNG
151 TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
251 ACCGTTACGA AGTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```

30 1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVFY RGT#WQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.*

35 *meningitidis*:

```

                                     10      20      30      40      50
orf13.pep      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
                                     |||||||
40 orf13a      MTVWFVA AVAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
                                     10      20      30      40      50      60
                                     60      70      80      90     100     110
orf13.pep      VHAKTAVRKVETDSYQDLGAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA
                                     |||||||
45 orf13a      VHAKTAVGKVETDSYQDLGAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                                     70      80      90      100     110     120
                                     120
orf13.pep      LIVRKEGNLLIITHPX
                                     |||||||
50

```

orf13a LIVRKEGNLLIIAKPX
130

The complete length ORF13a nucleotide sequence <SEQ ID 63> is:

```

5      1  ATGACTGTAT GGTGTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
      51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
     101  GCATTGCTTA CGGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
     151  GCTCTGCTTT CCGCGCTGGG TATTGTTTC GTACACGCCA AAACCGCCGT
     201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
     251  CCGAAATCCT CCGGCACGCA GCGGCAACC GTTACGAAGT TTTTATCGC
    10   301  GGTACGCACT GGCAGGCTCA AAATACGGGG CAAGAAGAGC TTGAACCAGG
     351  AACGCGCGCC CTAATCGTCC GCAAGGAAGG CAACCTTCTT ATCATCGCAA
     401  AACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 64>:

```

15      1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTA
      51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFYR
     101  GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP*

```

ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

```

20      10      20      30      40      50      60
orf13a.pep  MTVWFVAAVAVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
orf13-1      AVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
              10      20      30      40      50

25      70      80      90      100     110     120
orf13a.pep  VHAKTAVGKVETDSYQDLDAAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
orf13-1      VHAKTAVRKVETDSYQDLDAAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
              60      70      80      90      100     110

30      130
orf13a.pep  LIVRKEGNLLIIAKPX
orf13-1      LIVRKEGNLLIIHPX
              120

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N. gonorrhoeae*:

```

40      orf13      AVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF      51
      orf13ng      MTVWFVAAVAVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF      60

      orf13      VHAKTAVRKVETDSYQDLDAAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA      111
      orf13ng      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA      120

      orf13      LIVRKEGNLLIIHP      126
      orf13ng      LIVRKEGNLLIIANP      135

```

50 The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

```

55      1  ATGACTGTAT GGTGTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
      51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
     101  GCATTGCCTA CGGGCTGACT GGCAGCACGC CTGCCGCCGT CTTGACCGCC
     151  GCACTGCTTT CCGCGCTGGG CATTGTTTC GTACATGCCA AAACCGCCGT
     201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATACC GGAATATG
     251  CCGAAATCCT CCGATACACA GCGGCAACC GTTACGAAGT TTTTATCGC
     301  GGTACGCACT GGCAGGCGCA AAATACGGGG CAGGAAGTGT TTGAACCGGG
     351  AACGCGCGCC CTCATCGTCC GCAAAGAAGG TAACCTTCTT ATCATCGCAA
     401  ACCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVSAAL AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKAEILRYT GGNRYEVFYR
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

10 orf13-1.pep      10      20      30      40      50
      AVLIIELLTGT VYLLVSAALAGSGIAYGLTGSTPAAVLT XALLSALGIXF
orf13ng      10      20      30      40      50      60
      MTVWFVAAVA VLIIELLTGT VYLLVSAALAGSGIAYGLTGSTPAAVLTA ALLSALGIWF

15 orf13-1.pep      60      70      80      90     100     110
      VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVFYRGTHWQAQNTG QEELEPGTRA
orf13ng      60      70      80      90     100     110     120
      VHAKTAVGKVETDSYQDLDT GKAEILRYTGGNRYEVFYRGTHWQAQNTG QEVFEPGTRA

20 orf13-1.pep      120
      LIVRKEGNLLIITHPX
orf13ng      120
      LIVRKEGNLLIIANPX
      130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30 1  ATGTWTGATT TCGGTTTGG CGArCTGGTT TTTGTCTGGCA TTATCGCCCT
    51  GATWgtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
    101 GGcTCATCGG CAGGCTGCAA CGCTTTGTCTG GcAGCGTCAA ACAGGAATTT
    151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
    201 AGCTGCCGcC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
35 251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
    301 CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCTGATG AAAACGGCAA
    351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TCCGACGTT
    401 ATGCCGTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40 1  MXDFGLGELV FVGIIALIVL GPERXPEAR XAGRLIGRLQ RFVGSVKQEF
    51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
    101 LPEQRTPADF GVDENGNPXS RCGKHPIRRH FRRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45 1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCTGGCA TTATCGCCCT
    51  GATTGTCTCT GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
    101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
    151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
    201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
50 251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
    301 CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCTGATG AAAACGGCAA
    351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
    401 TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCCTTG GGACAGCGGG
    451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
    501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG

```

-95-

551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT
 601 TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTTC GTCCGAAACA
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

5 1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNLP DAANTLSDGI SDVMPSESY ASAETLGDGSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVPHTT
 201 SLRKQAISRK RDFRPKHRAK PKLRVRKS*

10 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTGCGCA TTATCGCCCT
 51 GATTGTCCTC GGCCCCGAAC GCCTGCCCCG GCGCGCCGCG ACCGCGCGAC
 101 GGCTCATCGG CAGGCTGCAA CGCTTTGTGCG GCAGCGTCAA ACAGGAATTT
 151 GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA
 201 AGCTGCCGCT GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
 251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCCCAAC AGCGCACGCC TGCTGATTTC GGTGTCGATG AAAACGGCAA
 351 TCCCTTTCCC GATGCGGCAA ACACCCTATT AGACGGCATT TCCGACGTTA
 401 TGCCGTCCGA ACCTTCCTAC GCTTCCGCGG AAACCCCTGG GGACAGCGGG
 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG
 501 GCGGGAATAC CTGACTGCTT CTGCGCGCGC ACCCGTCGTA CAGACGCTCG
 551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CCCCTGTTCC GCATACCACT
 601 TCGCTGCGTA AACAGGCAAT AAGCCGCAAA CGCGATTTGC GTCCTAAATC
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

25 This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNFP DAANTLLDGI SDVMPSESY ASAETLGDGSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVPHTT
 30 201 SLRKQAISRK RDLRPKSRK PKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

		10	20	30	40	50	60
35	orf2.pep	MXDFGLGELV	FVGIIALIVL	GPERXPEAAR	XAGRLIGRLQ	RFVGSVKQEF	DTQIELEELR
	orf2a	MFDFGLGELV	FVGIIALIVL	GPERLPEAAR	TAGRLIGRLQ	RFVGSVKQEF	DTQIELEELR
		10	20	30	40	50	60
40	orf2.pep	KAKQEFEEAAA	AQVRDSLKET	GTDMEGNLHD	ISDGLKPWEK	LPEQRTPADF	GVDENGPNXS
	orf2a	KAKQEFEEAAA	AQVRDSLKET	GTDMEGNLHD	ISDGLKPWEK	LPEQRTPADF	GVDENGPNFP
		70	80	90	100	110	120
45	orf2.pep	RCGKHPIRR	HFRRYAV				
	orf2a	DAANTLLDG	ISDVMPSESY	ASAETLGDG	SGQTGSTAEP	AEQDRAWREY	LTASAAAPVV
		130	140	150	160	170	180

50 The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

	orf2a.pep	MFDFGLGELV	FVGIIALIVL	GPERLPEAAR	TAGRLIGRLQ	RFVGSVKQEF	DTQIELEELR	60
	orf2-1	MFDFGLGELV	FVGIIALIVL	GPERLPEAAR	TAGRLIGRLQ	RFVGSVKQEF	DTQIELEELR	60
55	orf2a.pep	KAKQEFEEAAA	AQVRDSLKET	GTDMEGNLHD	ISDGLKPWEK	LPEQRTPADF	GVDENGPNFP	120
	orf2-1	KAKQEFEEAAA	AQVRDSLKET	GTDMEGNLHD	ISDGLKPWEK	LPEQRTPADF	GVDENGPNLP	120
60	orf2a.pep	DAANTLLDG	ISDVMPSESY	ASAETLGDG	SGQTGSTAEP	AEQDRAWREY	LTASAAAPVV	180

orf2-1 DAANTLSDGISDVMPERSYASAETLGDSGQTGSTAETDQDRAWREYLTASAAAPVV 180
 orf2a.pep QTVEVSYIDTAVETPVPHHTSLRKQAISRKRDLRPKSRAPKLRVRKSX 229
 5 orf2-1 QTVEVSYIDTAVETPVPHHTSLRKQAISRKRDLRPKSRAPKLRVRKSX 229

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 10 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQNSLHD ISDGLKPWEK
 101 LPEQRTPADF GVDEKGNLS RYKHRIRRH FRRYAV*

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTCGGCA TTATCGCCCT
 15 51 GATTGTCCTT GGTCCAGAAC GCCTGCCCCG AGCCGCCCCG ACTGCCGGAC
 101 GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT
 151 GACACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCTGA
 201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GATACGGATA
 251 TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCCGAAC AGCGCACGCG tgcgatttc gGTGTCGATg AAAacggcaa
 20 351 tccccctccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA
 401 TGCCGTCTGA ACGTTCCGAT ACTtccgcCG AAACCCCTTG GGACGACAGG
 451 CAAACCGGCA GTACAGCCGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG
 501 GCGGGAATAC CTGactgctt ctgcccgcgc acctgtcgta Cagagggccg
 551 tcgaagtcag ctaTATCGAT ACTGCTGTTG AAacgcctgT tccgcaCacc
 25 601 acttccctgc gcaAACAGGC AATAAACCGC AAACGCGATT Tttgtccgaa
 651 ACACCGCGCc aAACCGAAat tgcgcgtcCG TAAATCATAA

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 30 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQNSLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGNPLP DTANTVSDGI SDVMPERSD TSAETLGDDR
 151 QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETPVPH
 201 TSLRKQAINR KRDFCPKHRA KPKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

35 orf2.pep MXDFGLGELVFVGIIALIVLGPXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
 orf2ng MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
 40 orf2.pep KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS 120
 orf2ng KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNLSP 120
 orf2.pep RCGKHPIRRHFRRYAV 136
 45 orf2ng RYKHRIRRHFRYAV 136

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

50 orf2-1.pep MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
 orf2ng-1 MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
 55 orf2-1.pep KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP
 orf2ng-1 KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP

-97-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf2-1.pep	DAANTLSDGISDVMP	SERSYASAETLGD	SGQTGSTAEP	ETDQDRAWREY	LTASAAAPV	
	orf2ng-1	DTANTVSDGISDVMP	SERSDTSAETLGD	DRQTGSTAEP	ETDKDRAWREY	LTASAAAPV	
		130	140	150	160	170	180
10	orf2-1.pep	Q-TVEVSYIDTAVET	PVPHTTSLRKQ	AI	SRKRD	FRPKHRAKPKLR	VVRKSX
	orf2ng-1	QRAVEVSYIDTAVET	PVPHTTSLRKQ	AI	NRKRD	FCPKHRAKPKLR	VVRKSX
		190	200	210	220	230	

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined),
 and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein
 of *E.coli*:

```

gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
Score = 56.6 bits (134), Expect = 1e-07
Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

Query: 1  MFD FGLGELIFVGI IALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
          MFD G EL+ V II L+VLGP+RLP A +T I L+ +V+ EL +++L+E +
Sbjct: 1  MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60

Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87
          +K+ +A+ + LK + +++ +
Sbjct: 61 DSLKKVEKASLTNLTPELKASMDLRQA 88
  
```

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane
 proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above.
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A
 shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results
 of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice,
 whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis
 (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is
 a useful immunogen.

Example 10

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

```

40      1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
      51  CGC.TGCGGG AACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT
     101  TTgCGGTGCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
     151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
     201  CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGCTACTCC
     251  ATTGATGCAC kGrTwCstGG CGAATACATA AACAGCCCTG CCGTCCGTAC
     301  CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
     351  GTTTGACAGG TTTAACCCT TCTTTATCTA CACTTAATGC CCCTGCACTC
     401  TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
     451  TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC
  
```

501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC
 551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
 601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCAGAAATG..

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

5 1 MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT
 101 DYTYPREYET AETTSAGLTG LTTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSF ANADTDVFIN
 201 IDVFGTIRNR TEM..

10 Further work revealed the complete nucleotide sequence <SEQ ID 79>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 15 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 20 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTT GCAGTAGACA
 25 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
 801 AGGAATTAAA CCGACGAAG GATTAATGGT CGATTTCTCC GATATCCGAC
 851 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCAGC AACATAGACA
 30 AGGACAACCT TGA

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPREYET AETTSAGLTG LTTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSF ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKNFAFEA
 251 AYKENYALWM GPYKVSQGIK PTEGLMVDFS DIRPYGNHTG NSAPSVEADN
 301 SHEGYGSDE VVRQHRQGP *

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

40 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 45 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 50 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTT GCAGTAGACA
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 55 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
 801 AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTTCTCC GATATCCAAC
 851 CATACGGCAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCAGC GACATAGACA
 951 AGGGCAACCT TGA

60 This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK

-99-

51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPYRETT AETTSGGLTG LTTSLSLTNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSQGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN
 301 SHEGYGSDE AVRRHRQGP *

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

10	orf15.pep	10	20	30	40	50	60
		<u>MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARA</u>					
	orf15a	<u>MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA</u>					
		10	20	30	40	50	60
15	orf15.pep	70	80	90	100	110	120
		KVALYIATMGDQSGSLTGGGRYSIDAXXXGEYINSPAVRTDYTPRYETTAETTSGGLTG					
	orf15a	KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG					
		70	80	90	100	110	120
20	orf15.pep	130	140	150	160	170	180
		LTTSLSLTNAPALSRQSDGSGSKSSLGLNIGGMGDYRNETLTNNPRDTAFLSHLVQTVF					
	orf15a	LTTSLSLTNAPALSRQSDGSGSKSSLGLNIGGMGDYRNETLTNNPRDTAFLSHLVQTVF					
25		130	140	150	160	170	180
	orf15.pep	190	200	210			
		FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM					
30	orf15a	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL					
		190	200	210	220	230	240

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

35	orf15a.pep	10	20	30	40	50	60
		<u>MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA</u>					
	orf15-1	<u>MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA</u>					
		10	20	30	40	50	60
40	orf15a.pep	70	80	90	100	110	120
		KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG					
	orf15-1	KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG					
		70	80	90	100	110	120
45	orf15a.pep	130	140	150	160	170	180
		LTTSLSLTNAPALSRQSDGSGSKSSLGLNIGGMGDYRNETLTNNPRDTAFLSHLVQTVF					
	orf15-1	LTTSLSLTNAPALSRQSDGSGSKSSLGLNIGGMGDYRNETLTNNPRDTAFLSHLVQTVF					
		130	140	150	160	170	180
50	orf15a.pep	190	200	210	220	230	240
		FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL					
	orf15-1	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL					
55		190	200	210	220	230	240
	orf15a.pep	250	260	270	280	290	300
		IKPKTNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFS					
60	orf15-1	IKPKTNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFS					
		250	260	270	280	290	300
65	orf15a.pep	310	320				
		SHEGYGSDEAVRRHRQGQXPX					
	orf15-1	SHEGYGSDEVVRQHRQGQXPX					

310

320

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

```

      1  ATGCGGGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
      51  CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGCAAACGCT
5    101  TCGCGGTGCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
      151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
      201  AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
      251  TTGATGCACT GATTCGCGGC GAATACATAA ACAGCCCTGC CGTCCGCACC
      301  GATTACACCT ATCCGCGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
      351  TTTGACGGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
      401  CGCGCACCCA ATCAGACGGT AGCGGAAGTA GGAGCAGTCT GGGCTTAAAT
      451  ATTGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CCAACCGCG
      501  CGACACTGCC TTTCTTTCCC ACTTGGTGCA GACCGTATTT TTCCTGCGCG
      551  GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
      601  ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
      651  TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTT GCAGTAGACA
      701  GAACCAATAA AAAATTGCTC ATCAAACCCA AAACCAATGC GTTTGAAGCT
      751  GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
      801  AGGAATCAAA CCGACGGAAG GATTGATGGT CGATTCTCTC GATATCCAAC
      851  CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
      901  AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCAC AACATAGACA
      951  AGGGCAACCT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

```

      1  MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
      51  DMDLQALHGR KVALYIATMG DQSGSGLTGG RYSIDALIRG EYINSPAVRT
      101  DYTYPYRSET AETTSGLTGT LTSLSTLNA PALSRQSDG SGRSSLGLN
      151  IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
      201  IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
      251  AYKENYALWM GPYKVSQGIK PTEGLMVDFA DIQPYGNHTG NSAPSVEADN
      301  SHEGYGYSDE AVRQHRQGP

```

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

```

      orf15.pep  MQARLLIPILFSVFILSACGLTLTGIPSHGGGKRFQVEQELVAASARAAVKMDLQALHGR  60
      35  orf15ng    MRARLLIPILFSVFILSACGLTLTGIPSHGGGKRFQVEQELVAASARAAVKMDLQALHGR  60

      orf15.pep  KVALYIATMGDQSGSGLTGGRYSIDAXXGEYINSPAVRTDYTYPRYETTAETTSGLTGT  120
      40  orf15ng    KVALYIATMGDQSGSGLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTGT  120

      orf15.pep  LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF  180
      orf15ng    LTSLSTLNAPALSRQSDGSGSRSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF  180

      45  orf15.pep  FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM  213
      orf15ng    FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL  240

```

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

```

      50  orf15-1.pep  MQARLLIPILFSVFILSACGLTLTGIPSHGGGKRFQVEQELVAASARAAVKMDLQALHGR
      orf15ng    MRARLLIPILFSVFILSACGLTLTGIPSHGGGKRFQVEQELVAASARAAVKMDLQALHGR

      55  orf15-1.pep  KVALYIATMGDQSGSGLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTGT
      orf15ng    KVALYIATMGDQSGSGLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTGT

      60  orf15-1.pep  LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF

```

	orf15ng	: LTTSLSLTNAPALSRTQSDGSGSRSSLGLNIGMGDYRNETLTNPRDTAFSLHLVQTVF	130	140	150	160	170	180
5	orf15-1.pep	190 200 210 220 230 240 FLRGIDVVS PANADTDVFINIDVFGTIRNRTMHLNAETLKAQTKLEYFAVDRTNKKLL						
	orf15ng	: FLRGIDVVS PANADTDVFINIDVFGTIRNRTMHLNAETLKAQTKLEYFAVDRTNKKLL	190	200	210	220	230	240
10	orf15-1.pep	250 260 270 280 290 300 IKPKTNAFEAAAYKENYALWMGPYKVS KGIKPT EGLMVD FSDIRPYGNHTGNSAPSVEADN						
	orf15ng	: IKPKTNAFEAAAYKENYALWMGPYKVS KGIKPT EGLMVD FSDIQPYGNHTGNSAPSVEADN	250	260	270	280	290	300
15	orf15-1.pep	310 320 SHEGYGYSDEVVRQHRQGQPX						
	orf15ng	: SHEGYGYSDEAVRQHRQGQPX	310	320				

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

35	1	..GG.CAGCACA	AAAAACAGGC	GGTTGAACGG	AAAAACCGTA	TTTACGATGA
	51	TGCCGGGTAT	GATATTCGGC	GTATTCACGG	GCGATTCTC	CGAAAATAT
	101	ATCCCCGCGT	TCGGGCTTCA	AATTTTCTTC	ATCCTGTTT	TAACCCGCGT
	151	CGCATTCAAA	ACACTGCATA	CCGACCTTCA	GACCGCATCC	CGCCCGCTGC
40	201	CCGGACTGCC	CrGACTGACT	GCGGTTTCCA	CACTGTTCCG	CACAATGTCG
	251	AGCTGGGTCG	GCATAGGCGG	CGGTTCACTT	TCCGTCCCTC	TCCTAATCCA
	301	CTGCGGCTTC	CCCCGCCATA	AAGCCATCGG	CACATCATCC	GGCCTTGCCT
	351	GGCCGATTGC	ACTCTCCGGC	GCAATATCGT	ATCTCTCAA	CGGCTTGAAT
45	401	ATTGCAGGAT	TGCCCGAAGG	GTCACTGGGC	TTCTTTTACC	TGCCCGCCGT
	451	CGCCGTCCTC	AGCCCGGCCA	CCATTGCTTT	TGCCCGCTCT	GGTGTCAAAA
	501	CGGCCACAAA	ACTTTCTTCT	GCCAAACTCA	AAAAATC.TT	CGGCATTATG
	551	TTGCTTTTGA	TTGCCGGAAT	AATGCTGTAC	AACCTGCTTT	AA

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

1 ..GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILFLTAV
51 AFKTLHTDPO TASRPLPLGP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH

5	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGCGGC
	51	AGGTTTTTAT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTCTGTTT	ATGGTGCTTT	GATTTCGAGG	GTTTGGCACA	ACATCCTTAC
	151	CGCGAACACT	TCGCGCTCGG	CACATCCTTC	GCCGTCATGG	TCCTCACCGC
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGAAAA
10	251	CCGTATTTAC	GATGATGCCG	GGTATGATAT	TCGGCGTATT	CACGGGCGCA
	301	CTCTCCGCAA	AATATATCCC	CGCGTTCGGG	CTCCAAATTT	TCTTCATCCT
	351	GTTTTTAAAC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG
	451	TTCCGGCACA	TGTCGAGCTG	GGTCGGCATA	GCGCGCGGTT	CACTTTCCGT
15	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTT
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAAAGGTCAC	TGGGCTTCTG
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCAAT	CGCTTTGCCC
	701	CGCTCGGTGT	CAAAACGCCC	CACAAACTTT	CTTCTGCCAA	ACTCAAAAAA
20	751	Tc . TTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTTAA				

25

1	<u>MWHWDIILIL</u>	<u>LAVGSAAGFI</u>	<u>AGLFGVGGGT</u>	<u>LIVPVVLWVL</u>	<u>DLOGLAQHPY</u>
51	<u>AQHLAGVTSF</u>	<u>AVMVTFATFS</u>	<u>MLGQHKQAV</u>	<u>DWKTVFTMPG</u>	<u>GMIFGVFTGA</u>
101	<u>LSAKYIPAFG</u>	<u>LQIFFILFLT</u>	<u>AVAFKTLHTD</u>	<u>PQTASRPLPG</u>	<u>LPGLTAVSTL</u>
151	<u>FGTMSSWVGI</u>	<u>GGGSLVFPFL</u>	<u>IHCGFPAHKA</u>	<u>IGTSSGLAWP</u>	<u>IALSGAISYL</u>
201	<u>LNGLNIAGLP</u>	<u>EGSLGFLYLP</u>	<u>AVAVLSAATI</u>	<u>AFAPLGVKTA</u>	<u>HKLSSAKLKK</u>
251	<u>XFGIMLLLIA</u>	<u>GKMLYNLL*</u>			

	ORF17	3	HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP	59
			HK + + V + P ++ VF G F + +IF +++L ++ D	
	HI0902	72	HKLGNIWVQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD-	130
35	ORF17	60	QTASRPLPGLPXLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPI	119
			Q ++ L L + L G SS GIGGG VPFL G +AIG+S+ +	
	HI0902	131	QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMML	189
40	ORF17	120	ALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVXXXXXXXXXXXXXXXXX	179
			+SG S+++G +PE SLG++YLPV ++A.+ + LG	
	HI0902	190	GISGMFSFIVSGWGNPLMPEYSLGYIYLPVVGITATSFFTSKLGASATAKLPVSTLKKG	249
	ORF17	180	FGIMLLLIAGKM	191
			F + L+++A M	
45	HI0902	250	FAFLIVVAINM	261

```

50                                     10      20      30
    orf17.pep                        GQHKKQAVNGKTVFTMMPGMIFGVFTGAFS
                                     |||||:|||||:|||||:|:|
    orf17a      QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMVFGVFAGALS
                50      60      70      80      90      100

55                                     40      50      60      70      80      90
    orf17.pep      AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRLPGLPXLTA VSTLFGTMSSWVGIGG
                  |||||:|||||:|||||:|||||:|||||:|||||:|:|
    orf17a      AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRLPGLPGLTA VSTLFGTMSSWVGIGG

```

-103-

		110	120	130	140	150	160
5	orf17.pep	100	110	120	130	140	150
		GSLSVFPLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	AV				
	orf17a	170	180	190	200	210	220
		GSLSVFPLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	AV				
10	orf17.pep	160	170	180	190		
		AVLSAATIAFAPLGVKTAHKLSSAKLKKSF	GIMLLLIAGKMLYNLLX				
	orf17a	230	240	250	260		
		AVLSAATIAFAPLGVKTAHKLSSAKLKKSF	GIMLLLIAGKMLYNLLX				

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

15	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGC	GGC
	51	AGGTTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC	
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC	
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTCATGG	TCTTCACCGC	
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGA	AAA
20	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA	
	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCATCCT	
	351	GTTTTTAACC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	CCTCAGACCG	
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG	
	451	TTCCGACCAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACCTTCCGT	
25	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT	
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG	
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAAGGGTCAC	TGGGCTTCCT	
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCC	
	701	CGCTCGGTGT	CAAAACCGCC	CACAACTTT	CTTCTGCCAA	ACTCAAAAA	
30	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT	
	801	GCTTTAA					

This encodes a protein having amino acid sequence <SEQ ID 90>:

	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY	
	51	AQHLAVGTSF	AVMVETAFFS	MLGQHKQAV	DWKTVFTMMP	GMVFGVFAGA	
35	101	LSAKYIPAFG	LQIFFILFLT	AVAFKTLHTD	PQTASRPLPG	LPGLTAVSTL	
	151	FGTMSSWVG	GGSLSVFPL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL	
	201	LNLNLIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKK	
	251	SFGIMLLLIA	GKMLYNLL*				

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

40	orf17a.pep	10	20	30	40	50	60
		MWHWDIILILLAVGSAAGFIAGLFGVGGGT	LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF				
	orf17-1	10	20	30	40	50	60
		MWHWDIILILLAVGSAAGFIAGLFGVGGGT	LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF				
45	orf17a.pep	70	80	90	100	110	120
		AVMVETAFFSSMLGQHKQAVDWKTFTMMPGMVFGVFAGALSAKYIPAFGLQIFFILFLT					
	orf17-1	70	80	90	100	110	120
		AVMVETAFFSSMLGQHKQAVDWKTFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT					
50	orf17a.pep	130	140	150	160	170	180
		AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA					
55	orf17-1	130	140	150	160	170	180
		AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA					
60	orf17a.pep	190	200	210	220	230	240
		IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	PAVAVLSAATIAFAPLGVKTA				
	orf17-1	190	200	210	220	230	240
		IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	PAVAVLSAATIAFAPLGVKTA				
65	orf17a.pep	250	260	269			
		HKLSSAKLKKSF	GIMLLLIAGKMLYNLLX				

```

      ||||| |||||
orf17-1  HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
           250       260

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from *N. gonorrhoeae*:

```

      orf17.pep                                GQHKKQAVNGKTVFTMMPGMI FGVFTGAFS      30
      |||||: ||:|||||||:|||||
10  orf17ng      QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKKQAVDWKTI FAMMPGMI FGVFAGALS      102
      orf17.pep      AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGMTSSWVGIGG      90
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15  orf17ng      AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTA VSTLFGAMSSWVGIGG      162
      orf17.pep      GSLSVPFLIHC GFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV      150
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf17ng      GSLSVPFLIHC GFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPVAV      202
20  orf17.pep      AVLSAATIAFAPLGVKTA HKLSSAKLKKSF GIMLLLIAGKMLYNLL      196
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf17ng      AVLSAATIAFAPLGVKTA HKLSSAKLKESFGIMLLLIAGKMLYNLL      268

```

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

```

25      1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
      51  AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTI FAMMP GMIFGVFAGA
      101  LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
      151  FGAMSSWVGI GGSLSVPFL IHC GFPAHKA IGTSSGLAWP IALSGAISYL
      201  VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
30  251  SFGIMLLLIA GKMLYNLL*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

```

      1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCcgtag gcAGTGCGGC
      51  AGGTTTTATT GCCGGCCTGT Tcgggttagg cggcgGTACG CTGATTGTCC
35  101  CTGTCGTTTT ATGGGTGCTT GATTTCAGG GTTTGGCACA ACATCCTTAC
      151  GCGCAACACC TCGCCGTCGG CAcaTccttc gcCGTCATGG TCTTCACCGC
      201  CTTTTCCAGT ATGTTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
      251  CCATATTTGC GATGATGCCG GGTATGATAT TCGGCGTATT CGCTGGCGCA
      301  CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
      351  GTTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGGT CGTCAGACGG
40  401  CATCCCGCCC GCTGCCCGGG CTGCCCGGAC TGACTGCGGT TTCCACACTG
      451  TTCGGCGCAA TGTCGAGCTG GGTCCGCATA GGCGGCGGTT CACTTTCCGT
      501  CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
      551  CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
      601  GTCAACGGTC TGAATATTGC AGGATTGCCC GAAGGTCGC TGGGCTTCCT
45  651  TTACCTGCCC GCCGTCGCGC TCCTCAGCGC GGCAACCATT GCCTTTGCC
      701  CGCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAGAA
      751  TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801  GCTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

```

50      1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
      51  AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTI FAMMP GMIFGVFAGA
      101  LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
      151  FGAMSSWVGI GGSLSVPFL IHC GFPAHKA IGTSSGLAWP IALSGAISYL
55  201  VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
      251  SFGIMLLLIA GKMLYNLL*

```

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

```

           10       20       30       40       50       60
orf17-1.pep  MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF

```

-105-

	orfl7ng-1	 MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
		10 20 30 40 50 60
5	orfl7-1.pep	70 80 90 100 110 120 AVMVFTAFSSMLGQHKKQAVDWKTIVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
	orfl7ng-1	 AVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVFAGALSAKYIPAFGLQIFFILFLT
10		70 80 90 100 110 120
	orfl7-1.pep	130 140 150 160 170 180 AVAFKTLHTDPQTASRLPGLPGLTAVSTLFGTMSSWVGIGGSLSVPFLLHCGFPAHKA
	orfl7ng-1	 AVAFKTLHTGRQTASRLPGLPGLTAVSTLFGAMSSWVGIGGSLSVPFLLHCGFPAHKA
15		130 140 150 160 170 180
	orfl7-1.pep	190 200 210 220 230 240 IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA
20	orfl7ng-1	 IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA
		190 200 210 220 230 240
	orfl7-1.pep	250 260 269 HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
25	orfl7ng-1	 HKLSSAKLKESFGIMLLLIAGKMLYNLLX
		250 260

In addition, ORF17ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

30 sp|P44070|Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir||G64015 hypothetical protein
HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae
predicted coding region HI0902 [Haemophilus influenzae]Length = 264
Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
Identities = 15/43 (34%), Positives = 23/43 (53%)

35 Query: 55 AVGTSFAVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVF 97
A+GTSFA +V T S HK + W+ + + P ++ VF
Sbjct: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94
Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
Identities = 44/114 (38%), Positives = 65/114 (57%)

40 Query: 150 LFGAMSSWVGIGGSLSVPFLLHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209
L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G +
45 Sbjct: 148 LIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207
Query: 210 PEGSLGFLYLPAAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263
PE SLG++YLPVAV ++A + + LG KL + LK+ F + L+++A M
Sbjct: 208 PEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261

50

This analysis, including the homology with the hypothetical *H. influenzae* transmembrane protein, suggests that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 12

55 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 95>:

60 1 ..GGAAACGGAT GGCAGGCAGA CCCCGAACAT CCGCTGCTCG GGCTTTTTCG
51 CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCCGGAATA TGTGCGTTGG
101 TGCATTATTG CTTTTCCGGA ACGGTTCAAG TGTGTGTGT TGCGGCACTG
151 CTCAAACCTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
201 GCTGATGGCG GTTGCTATG TCCACCGCTG CCGTATAGAC CGGCAGCCGC
251 CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGGCAGCG

5 1 ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
51 LKLYALKPVY WFLVQFVLMA VAYVHRCGID RQPPSTFGGS QLRLGGLTAA
101 LMOVSVLVLL LSEIGR*

	1	ATGATTTTGC	TGCATTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
10	51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCCC
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TGTGTCATCC	CCCATTTTTA
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGCATTGG	AACCGGAAAA
	251	CAGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
15	301	TTTGCCGTC	GTAATGTATC	GATGACGCTT	GCTTTTGTGC	GAATATGTGC
	351	GTTGGTGAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT	GTTGTTGCGG
	401	CAC TGCTCAA	ACTTTATGCG	CTGAAGCCGG	TCTGTTGGTT	CGTGTTGCAG
	451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
	501	GCCGCCGTC	ACGTTCCGGC	GCTCGCAGCT	GCGACTCGGC	GGGTTGACGG
20	551	CAGCGTTGAT	GCAGGTCTCG	GTACTGGTGC	TGCTGCTTTC	AGAAATTGGA
	601	AGATAA				

25

1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP
51	GIWGMTRAAP	LFIPHFYLT	LSIFFFIQHW	NRKTDGNGWQ	ADPEHPLLGL
101	FAVSNVSMTL	AFVGICALVH	YCFSGTQVQF	VFAALLKLYA	LKPYYWVFLQ
151	FVLMAVAYVH	RCGIDRQPPS	TFGGSQRLRG	GLTAALMQVS	VLVLLLSEIG
201	R*				

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 *meningitidis:*

```

35 orf18.pep                                     GNGWQADPEHPLLGLFAVSNVSMTLAFVGI
                                     |||||
orf18a.    TRAAPLFIPHPFYLTLSIFFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI
           60          70          80          90          100          110

40 orf18.pep           40          50          60          70          80          90
           CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
           |||||
orf18a      CALVHYCFSXTQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
           120          130          140          150          160          170

45 orf18.pep           100          110
           QRLRGGLTAALMQVSVLVLLLSEIGRX
           |||||
orf18a      QRLRGGLTAALMQXSVLVLLLSEIGRX
           180          190          200

```

50	1	ATGATTTTGC	TGCATTTTGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
	51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCC
	151	GGCATATGGG	GAATGACCG	CGCCGCGCCC	TTGTTCATCC	CCCATTTTTC
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGCATATG	AACCGGAAAA
55	251	CGGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCTCT	GCTCGGGCTG
	301	TTTGCCGTC	GTAATGTATC	GATGACGCTT	GCTTTTGTCT	GAATATGTGC
	351	GTTGGTGTC	TATTGCTTTT	CGNGAACGGT	TCAAGTGTTT	GTGTTTGGCG
	401	CACCTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTGTGCA

-107-

```

451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501 GCCGCCGTCA ACGTTCGGCG GNTCGCAGCT GCGACTCGGC GGGTTGACGG
551 CAGCGTTGAT GCAGNTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
601 AGATAA

```

5 This encodes a protein having amino acid sequence <SEQ ID 100>:

```

1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
51 GIWGMTRAAP LFIPHFYLT LGSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSMTL AFGICALVH YCFSXTVQVF VFAALLKLYA LKPVYWFVLQ
151 FVLMAYAVVH RCGIDRQPPS TFGGSQRLRG GLTAALMQXS VLVLLLSEIG
201 R*

```

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

```

10 20 30 40 50 60
orf18a.pep MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
15 orf18-1 MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
10 20 30 40 50 60
20 70 80 90 100 110 120
orf18a.pep LFIPHFYLT LGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
20 orf18-1 LFIPHFYLT LGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
70 80 90 100 110 120
25 orf18a.pep YCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQRLRG
orf18-1 YCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQRLRG
130 140 150 160 170 180
30 orf18a.pep GLTAALMQXS VLVLLLSEIGRX
orf18-1 GLTAALMQXS VLVLLLSEIGRX
190 200
35

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N.gonorrhoeae*:

```

40 orf18.pep GNGWQADPEHPLLGLFAVSNVSMTLAFVGI 30
orf18ng TRAAPLFIPHFYLT LGSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI 115
orf18.pep CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS 90
45 orf18ng CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS 175
orf18.pep QLRGLGLTAALMQXS VLVLLLSEIGR 116
orf18ng QLRGLVLAAMLQVAVTAMLLAEIGR 201

```

50 The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

```

1 ATGATTTTGC TGCATTTGGA TTTTGTCT GCCTTACTGt aTGCGGcggt
51 tttTctgTTT CTGATATCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101 GTATTGCGTT GTGGCTCGGC ATCTCGGTTT TAGGGGTAA GCTGATGCCG
55 151 GGGATGTGGG GAATGACCG CGCCGCGCCT TTGTTTCATCC CCCATTTTAA
201 CCTGACTTTG GGCAGCATAT TTTTTCAT CGGGTATTGG AACCGGAAAA
251 CAGATGGAAA CGGATGGCAG GCAGACCCG AACATCCGCT GCTCGGGCTT
301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTGCG GAATATGTGC
351 GTTGGTGCAT TATTGCTTT CGGGAACGGT TCAAGTGTTC GTGTTGCGG
401 CATTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTCGAG
60 451 TTTGTATTGA TGGCGGttgC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501 GCCGCCGTCA ACGTTCGGCG GTTCGCAGCT GCGACTCGGC GTGTTGCGCG

```

-108-

551 CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC
601 AGATGA

This encodes a protein having amino acid sequence <SEQ ID 102>:

5 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLG ISVLGVKLMP
 51 GMWGMTRAAP LFIPHFYLT L GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL
 101 FAVSNVSM TL AFGVICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
 151 FVLMAVAYVH RCGIDRQPPS TFGGSQRLRG VLAAMLMOVA VTAMLLAEIG
 201 R*

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

10		10	20	30	40	50	60
	orf18-1.pep	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIALWLG	ISVLGVKLMP	GIWGMTRAAP
	orf18ng	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIALWLG	ISVLGVKLMP	GMWGMTRAAP
15		70	80	90	100	110	120
	orf18-1.pep	LFIPHFYLT	L GSIFFFIGH	WNRKTDGNGWQ	ADPEHPLLGL	FAVSNVSM TL	AFVGICALVH
	orf18ng	LFIPHFYLT	L GSIFFFIGY	WNRKTDGNGWQ	ADPEHPLLGL	FAVSNVSM TL	AFVGICALVH
20		70	80	90	100	110	120
	orf18-1.pep	YCFSGTVQVF	VFAALLKLYA	LKPVYWFVLQ	FVLMAVAYVH	RCGIDRQPPS	TFGGSQRLRG
	orf18ng	YCFSGTVQVF	VFAALLKLYA	LKPVYWFVLQ	FVLMAVAYVH	RCGIDRQPPS	TFGGSQRLRG
25		130	140	150	160	170	180
	orf18-1.pep	GLTAALMQV	SVLVLLSEIGRX				
	orf18ng	VLAAMLMOVA	VTAMLLAEIGRX				
30		190	200				

Based on this analysis, including the presence of several putative transmembrane domains in the
35 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

40 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT
 51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
 101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTTCGAT
 151 TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC
 201 CCTGTTACAC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
45 301 GGCGCGGNCG ...

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

50 1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNXXTGR LK NIITVALEF LSSLTAQSTL GTGLPFILAM TLMXXFTIL
 101 GAX...

Further work revealed the complete nucleotide sequence <SEQ ID 105>:

 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
 51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
 101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTTCGAT
 151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC

-109-

201 CCTGTTACACC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA
 301 GGGCGGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
 401 ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
 451 CTGTTCCAAA TCGTCTCGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
 501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
 551 ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
 651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
 701 GTTACTACTT TGCCGCCCAA GACATACAGC AACGCATCAG CTCCGCCAC
 751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAAACCCG ACATCATCTT
 801 CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
 851 CCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
 901 CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
 951 CGACAGTCCC GACATCCGCC ACCTGCGCGC CTTTCTCGAC AACCTCGGCA
 1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
 1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
 1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
 1151 TATTCGGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCCG CGCCTGCACC
 1201 ATCGTCGAAG CCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
 1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAGC CGCGTCCGCC
 1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
 1351 TACTTCACCC CGTCTGTGCA AACCAAACTC TGGATTGTCA TCGCCAGTAC
 1401 CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
 1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
 1501 TACGCGGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
 1551 TGCTCGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
 1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
 1701 CGCGGCCACC CGCCGCCCGC CCCACGAACA CACCGCGGCC CTCAGCAGCA
 1751 CCTTTTCGGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
 1801 CCCGGCTTTA CCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
 1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
 1901 TTACCGCACA GTTCCACCTC CGCGCCGAAC ACACCGCCA CATCTTCCAA
 1951 CACCTGCCCG AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
 2001 GCGGGCGGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGCT CGAACCTAC
 2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
 2151 A

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGRLL NIITVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
 151 LFIQIVLPHRP QVESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
 201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIFRIH RLLEMGGQAC RNTAQALRAS KDYVYSKRLG
 301 RAIEGCRQSL RLLSDSNDSP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
 351 NDRMGDTRIA ALETSSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAACCT
 401 IVEALNINLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
 451 YFTPSVETKL WIVIASTTLF FMTRYKYSF STFFITIQAL TSLSLAGLDV
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSNGAYLE
 551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
 601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAHTAHIFQ
 651 HLPETEPDDF QTALDTRLGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
 701 YRAYRQIPHR QPQNAA*

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

60 orf19 6 LKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLKNIITT 65
 L +I+++PVF +V AA +W +MP +LGIIAGGLVDLDN TGRLKN+ T
 YHFK 5 LNAKIVSTIPVFIADVIAAVGIWFFDISQSMPLILGIIAGGLVDLDNRLTGRLKNVFFT 64

orf19 66 VALFTLSSLTAQSTLTGTLPLFILAMTMTXXFTILGA 102
 + F++SS Q +G + +I+ MT++T FT++GA
 YHFK 65 LIAFSISSFIVQLHIGKPIQYIVLMTVLTFIFTMIGA 101

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
10	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLLK					
	orf19a	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLLK					
		10	20	30	40	50	60
15	orf19.pep	NIIITVALFTLSSLTAQSTLTGTLPLFILAMTMTXXFTILGAX					
	orf19a	NIIATVALFTLSSLVAQSTLTGTLPLFILAMTMTFGFTIMGAVGLKYRTFAFGALAVATY					
		70	80	90	100		
20	orf19a	TTLTYTPETYWLTNPFMLICGTVLYSTAILFQIILPHRPVQENVANAYEALGSSYLEAKA					
		130	140	150	160	170	180

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

	1	ATGAAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTG	GGCGAACCCA
25	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCTGGCGG	CCTGGTTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTCAAC	CTCTCCTCAC	TTGTGCGCGA	AAGCACCTC	GGCAGAGGTT
	251	TGCCATTTCAT	CCTCGCCATG	ACCCTGATGA	CTTTCGGCTT	TACCATCATG
	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
30	351	CGCCACCTAC	ACCACACTTA	CCTACACCCC	CGAACCTAC	TGGCTGACCA
	401	ACCCCTTTAT	GATTCTGTGC	GGAACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTTCAAGAAA	ACGTCGCCAA
	501	CGCCTACGAA	GCACTCGGCA	GCTACCTCGA	AGCCAAAGCC	GACTTTTTTCG
	551	ATCCCGACGA	AGCCGAATGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
35	601	AGCAACACCG	GCCTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGCCTT	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATACACG	AACGCATCAG	CTCCGCCCAT
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATT	AAAAACACCG	ACATCATCTT
	801	CCGCATCCAC	CGCCTGCTCG	AAATGCAGGG	ACAAGCCTGC	CGCAACACCG
40	851	CCCAAGCCCT	GCGCGCAAGC	AAAGACTACG	TTTACAGCAA	ACGCCTCGGC
	901	CGCGCCATCG	AAGGCTGCCG	CCAATCGCTG	CGCCTCCTTT	CAGACAGCAA
	951	CGACAATCCC	GACATCCGCC	ACCTGCGCCG	CCTTCTCGAC	AACCTCGGCA
	1001	GCGTCGACCA	GCAGTTCGCG	CAACTCCAGC	ACAACGGCCT	GCAGGCAGAA
	1051	AACGACCGCA	TGGGCGACAC	CCGCATCGCC	GCCCTCGAAA	CCGGCAGCCT
45	1101	CAAAAACACC	TGGCAGGCAA	TCCGTCCGCA	GCTAAACCTC	GAATCAGGCG
	1151	TATTCGGCCA	TGCCGTCCGC	CTGTCCCTTG	TCGTTGCCGC	CGCCTGCACC
	1201	ATCGTCGAAG	CCCTCAACCT	CAACCTCGGC	TACTGGATAC	TACTGACCGC
	1251	CCTTTTCGTC	TGCCAACCCA	ACTACACCGC	CACCAAAAGC	CGCGTCCGCC
	1301	AGCGCATCGC	CGGCACCGTA	CTCGGCGTAA	TCGTCGGCTC	GCTCGTCCCC
50	1351	TACTTTACCC	CCTCCGTCGA	AACCAAACTC	TGGATCGTCA	TCGCCAGTAC
	1401	CACCCTCTTT	TTTATGACCC	GCACCTACAA	ATACAGCTTC	TCGACATTTT
	1451	TCATCACCAT	TCAAGCCCTG	ACCAGCCTCT	CCCTCGCAGG	GTTGGACGTA
	1501	TACGCCGCCA	TGCCCGTAGC	CATCATCGAC	ACCATTATCG	TGCCTCCCTT
	1551	TGCCTGGGCG	GCACTCAGCT	ACCTGTGGCC	AGACTGGAAA	CAGCTCACGC
55	1601	TCGAACGCAC	CGCCGCCCTT	GCCGTATGCA	GCAACGGCGC	CTATCTCGAA
	1651	AAAATCACCG	AACGCCTCAA	AAGCGGCGAA	ACCGGCGACG	ACGTCGAATA
	1701	CCGCGCCACC	CGCCGCCGCG	CCCACGAACA	CACCGCCGCC	CTCAGCAGCA
	1751	CCCTTTCCGA	CATGAGCAGC	GAACCGGCAA	AATTGCGCGA	CAGCCTGCAA
	1801	CCCGGCTTTA	CCCTGCTCAA	AACCGGCTAC	GCCCTGACCG	GCTACATCTC
60	1851	CGCCCTCGGC	GCATACCGCA	GCGAAATGCA	CGAAGAATGC	AGCCCCGACT
	1901	TTACCGCACA	GTTCCACCTC	GCCGCGGAAC	ACACCGCCCA	CATCTTCCAA
	1951	CACCTGCCCG	AAACCGAACC	CGACGACTTT	CAGACAGCAC	TGGATACACT
	2001	GCGCGGCGAA	CTCGACACCC	TCCGCACCCA	CAGCAGCGGA	ACACAAAGCC
	2051	ACATCCTCCT	CCAACAGCTC	CAACTCATCG	CCCGGCAGCT	CGAACCCCTAC
65	2101	TACCGGCCTT	ACCGACAAAT	TCCGCACAGG	CAGCCCCAAA	ACGCAGCCTG
	2151	A				

This encodes a protein having amino acid sequence <SEQ ID 108>:

```

1  MKTPPLKPLL  ITSLPVFASV  FTAASIVWQL  GEPKLAMPFV  LGIIAGGLVD
51  LDNRLTGRLK  NIIATVALFT  LSSLVAQSTL  GTGLPFIAM  TLMTFGFTIM
101 GAVGLKYRTF  AFGALAVATY  TTLTYTPETY  WLTNPFMILC  GTVLYSTAIL
5  151  LFQIILPHRP  VQENVANAYE  ALGSYLEAKA  DFFDPDEAEW  IGNRHIDLAM
201  SNTGVITAFN  QCRSALFYRL  RGKRRHPRTA  KMLRYFFAAQ  DIHERISSAH
251  VDYQEMSEKF  KNTDIIIFRIH  RLLEMQGOAC  RNTAQALRAS  KDYVYSKRLG
301  RAIEGCRQSL  RLLSDSNDNP  DIRHLRRLLD  NLGSVDQOFR  QLQHNGLQAE
351  NDRMGDTRIA  ALETGSLKNT  WQAIRPOLNL  ESGVFRHAVR  LSLVVAAC
10  401  IVEALNLNLG  YWILLTALFV  CQPNYTATKS  RVRQRIAGTV  LGVIVGSLVP
451  YFTPSVETKL  WIVIASTTLF  FMTRTYKYSF  STFFITIQAL  TSLSLAGLDV
501  YAAMPVRIID  TIIGASLAWA  AVSYLWPDWK  YLTLERTAAL  AVCSNGAYLE
551  KITERLKSSE  TGDDVEYRAT  RRAHEHTAA  LSSTLSDMSS  EPAKFADSLQ
601  PGFTLLKGY  ALTGYISALG  AYRSEMHEEC  SPDFTAQFHL  AAHTAHIFQ
15  651  HLPETEPDDF  QTALDTLRGE  LDTLRTHSSG  TQSHILLQQL  QLIARQLEPY
701  YRAYRQIPHR  QPQNAA*

```

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

```

20  orf19a.pep      10      20      30      40      50      60
    MKTPPLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
    orf19-1        10      20      30      40      50      60
    MKTPPLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK

25  orf19a.pep      70      80      90      100     110     120
    NIIATVALFTLSSLVAQSTLGTGLPFIAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY
    orf19-1        70      80      90      100     110     120
    NIITVALFTLSSLTAQSTLGTGLPFIAMTLMTFGFTILGAVGLKYRTFAFGALAVATY

30  orf19a.pep     130     140     150     160     170     180
    TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA
    orf19-1        130     140     150     160     170     180
    TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIVLPHRPVQESVANAYDALGGYLEAKA

35  orf19a.pep     190     200     210     220     230     240
    DFFDPDEAEWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKRRHPRTAKMLRYFFAAQ
    orf19-1        190     200     210     220     230     240
    DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKRRHPRTAKMLRYFFAAQ

40  orf19a.pep     250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFNTDIIIFRIHRLLEMQGOACRNTAQALRASKDYVYSKRLG
    orf19-1        250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFNTDIIIFRIHRLLEMQGOACRNTAQALRASKDYVYSKRLG

45  orf19a.pep     310     320     330     340     350     360
    RAIEGCRQSLRLLSDSNDNPDIRHLRRLLDNLGSVDQOFRQLQHNGLQAENDRMGDTRIA
    orf19-1        310     320     330     340     350     360
    RAIEGCRQSLRLLSDSNDSPDIRHLRRLLDNLGSVDQOFRQLQHNGLQAENDRMGDTRIA

50  orf19a.pep     370     380     390     400     410     420
    ALETGSLKNTWQAIRPOLNLESGVFRHAVRLSLVVAACTIVEALNLNLGYWILLTALFV
    orf19-1        370     380     390     400     410     420
    ALETSSLKNTWQAIRPOLNLESGVFRHAVRLSLVVAACTIVEALNLNLGYWILLTALFV

55  orf19a.pep     430     440     450     460     470     480
    CQPNYTATKSRVRQRIAGTVLGIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
    orf19-1        430     440     450     460     470     480
    CQPNYTATKSRVRQRIAGTVLGIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF

60  orf19a.pep     490     500     510     520     530     540
    STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL
    orf19-1        490     500     510     520     530     540
    STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL

```

-112-

	orf19-1	STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL	490	500	510	520	530	540
5	orf19a.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
	orf19-1	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
10	orf19a.pep	PGFTLLKGTGYALTGYISALGAYRSEMHEECSPDFTAQFHLLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
	orf19-1	PGFTLLKGTGYALTGYISALGAYRSEMHEECSPDFTAQFHLLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
15	orf19a.pep	QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	
20	orf19-1	QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N. gonorrhoeae*:

25	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK	60
30	orf19.pep	NIITTVALLFTLSSSLTAQSTLGTGLPFILAMTLMTXFTILGAX	103
	orf19ng	NIIATVALFTLSSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIAGGLVD
	51	LDNRLTGRK	NIIATVALFT	LSSSLTAQSTL	GTGLPFILAM	TLMTFGFTIL
	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAI
	151	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA	DDFDPDEAAW	IGNRHIDLAM
	201	SNTGVITAFN	QCRSALFYRL	RGKHRHPERTA	KMLRYFFAAQ	DIHERISSAH
40	251	VDYQEMSEKF	KNTDIIIFRIR	RLLEMQGGAC	RNTAQAIRSG	KDYVYSKRLG
	301	RAIEGCRQSL	RLLSDGNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHSDSPAE
	351	NDRMGDTRIA	ALETGSFKNT	*		

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

45	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCGCA	AAGCACCTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
50	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTCCAAGAAA	CGCTGCCCCA
	501	TGCCTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
55	551	ACCCCGATGA	GGCAGCCTGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
60	801	CCGCATCCGC	CGCCTGCTCG	AAATGCAGGG	GCAGGCGTGC	CGCAACACCG
	851	CCCAAGCCAT	CCGGTCGGGC	AAAGACTAcg	tTTACAGCAA	ACGCCCTCGGA
	901	CGCGCCATcg	aaggctgCCG	CCAGTCGCTg	cgccctCCTTt	cagacggcaA
	951	CGACAGTCCC	GACATCCGCC	ACCTGAGccg	CCTTCTCGAC	AACCTCGgca

-113-

1001 GCGTcgacca gcagtTCcgc caactCCGAC ACAgecgactC CCCCCGcgaa
 1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCCtcgaaa ccggcagctT
 1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG
 1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
 1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC
 1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGATCC
 1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
 1351 TACTTCACCC CCTCCGTCGA AACCAAACTC TGGATTGTCA TCGCCGTAC
 1401 CACCCTGTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
 1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
 1501 TACGCCGCCA TGCCCGTGCG CATCATcgac ACCATTATCG GCGCATCCCT
 1551 TGCTTGGGCG GCGGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTACGCG
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA
 1651 AAAATTGCCG AACGCCTCAA AACCGGCGAA ACCGGCGACG ACATAGAATA
 1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
 1751 CCCTTTCCGA CATGGGACG GAACCCGCAA AATTCCGCCA CAGCCTGCAA
 1801 CCGGCTTTA CCTGTCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
 1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
 1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCCA CATCTTCCAA
 1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATCACT
 2001 GCGCGGCGAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CcggGCAACT CGAACCTAC
 2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
 2151 A

25 This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGRK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMFTFGFTIL
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
 151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
 201 SNTGVITAFN QCRSALFYRL RGKRRHPTA KMLRYFFAAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIIFRIR RLEMQGQAC RNTAQAIRSG KDYVYSKRLG
 301 RAIEGCROSL RLSDGNDSP DIRHLSRLLD NLGSVDQOFR QLRHSDSPA
 351 NDRMGDTRIA ALETGSFKNT WQAIRPOLNL ESCVFRHAVR LSLVVAAC
 401 IVEALNLLG YWILLTALFV CQPNYTATKS RVYQRIAGTV LGVIVGSLVP
 451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITIQL TSLSLAGLDV
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSSGTYLQ
 551 KIAERLKTGE TGDDIEYRIT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
 601 PGFTLLKTYG ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAHTAHIFQ
 651 HLPDMGPDDF QTALDTRGE LGTLRTRSSG TQSHILLQQL QLIARQLEPY
 701 YRAYRQIPHR QPQNAA*

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

		10	20	30	40	50	60
orf19-1.pep		MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD	LDNRLTGRK
45	orf19ng-1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD	LDNRLTGRK
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf19-1.pep	NIIATVALFT	LSSLTAQSTL	GTGLPFILAM	TLMFTFGFTIL	GAVGLKYRTF	AFGALAVATY
	orf19ng-1	NIIATVALFT	LSSLTAQSTL	GTGLPFILAM	TLMFTFGFTIL	GAVGLKYRTF	AFGALAVATY
		70	80	90	100	110	120
55	orf19-1.pep	TTLTYTPETY	WLTNPFMILC	GTVLYSTAIL	LFQIILPHRP	VQESVANAYD	ALGGYLEAKA
	orf19ng-1	TTLTYTPETY	WLTNPFMILC	GTVLYSTAIL	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA
		130	140	150	160	170	180
60	orf19-1.pep	DFFDPDEAAW	IGNRHIDLAM	SNTGVITAFN	QCRSALFYRL	RGKRRHPTA	KMLRYFFAAQ
	orf19ng-1	DFFDPDEAAW	IGNRHIDLAM	SNTGVITAFN	QCRSALFYRL	RGKRRHPTA	KMLRYFFAAQ
		190	200	210	220	230	240
65	orf19-1.pep	DIHERISSAH	VQEMSEKF	KNTDIIIFRI	RLLEMQGQAC	RNTAQALR	ASKDYVYSKRLG

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

50
55
60
65

BNSDOCID: <WO _9924578A2_1_>

Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQPGFTLL 606
 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFAD+ P
 Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADTCNPALPCS 300

5 Query: 607 KTGyALTGYISALGAYRSEMHEECSP 632
 K ALTGYISALG ++ + +P
 Sbjct: 301 KPATALTGYISALGHAACTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology
 10 with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```

15      1  ATGAATATGC  TGGGAGCTTT  GGCAAAAGTC  GGCAGCCTGA  CGATGGTGTG
      51  GCGCGTTTTG  GGATTTGTGC  GCGATACGGT  CATTGCGCGG  GCATTCGGCG
      101  CGGGTATGGC  GACGGATGCG  TTTTTTGTCT  CGTTCAAAC  GCCCAACCTG
      151  CTTGCGCCGC  TGTTTGCGGA  GGGGGCGTTT  GCCCAAGCGT  TTGTGCCGAT
      201  TTTGGCGGAA  TACAAGGAAA  CGCGTTCAAA  AGAGGCGG.  GAAGCCTTTA
      251  TCCGCCATGT  GCGGGGATG  CTGTCGTTTG  TACTGGTTAT  CGTTACCGCG
      301  CTGGGCATAC  TTGCCGCGCC  TTGGGTGATT  TATGTTTCCG  CACCCGAGTT
      351  TTGCCCAAGA  TGCCGACAAA  TTTCAGCTCT  CCATCGATTT  GCTGCGGATT
      401  ACGTTTCCTT  ATATATTATT  GATTTCCCTG  TCTTCATTG  TCGGCTCGGT
      451  ACTCAATTCT  TATCATAAGT  TCGGCATTCC  GCGGTTTACG  CCAC. GTTTC
      501  TGAACGTGTC  GTTTATCGTA  TTCGCGCTGT  TTTTCGTGCC  GTATTTTCGAT
      551  CCGCCCGTTA  CCGCGCyGGC  GTGGGCGGTC  TTTGTGCGCG  GCATTTTGCA
      601  ACTCGrmTTC  CAACTGCCCT  GGCTGGCGAA  ACTGGGCTTT  TTGAAACTGC
      651  CCAAACtGAG  TTTCAAAGAT  GCGGCGGTCA  ACCGCGTGAT  GAAACAGATG
      701  GCGCCTGCgA  TTTTgGGCGT  GAgCGTGGCG  CAGGTTTCTT  TGGTGATCAA
      751  CACGATTTTc  GCGTCTTATC  TGCAATCGGG  CAGCGTTTCA  TGGATGTATT
      801  ACGCCGACCG  CATGATGGAG  CTGCCAGCG  GCGTGCTGGG  GCGGCGACTC
      851  GGTACGATTT  TGCTGCCGAC  TTTGTCCAAA  CACTCGGCAA  ACCaAGATAC
      901  GGaACAGTTT  TCCGCCCTGC  TCGACTGGGG  TTTGCGCCTG  TGCATGCTgc
      951  TGACGCTGCC  GGCGgcGGTC  GGACTGGCGG  TGTGTGCTGT  cCCgCtGGTG
      1001  GCGACGCTGT  TTATGTACCG  CGWATTTACG  CTGTTTGACG  CGCAGATGAC
      1051  GCAACACGCG  CTGATTGCCT  ATTCTTTCCG  TTTAATCGGC  TTAATCATGA
      1101  TTAAAGTGTT  GGCAACCGGC  TTCTATGCGC  GGCAAAACAT  CAAwAmGCCC
      1151  GTCAAATCG  CCATCTTAC  GCTCATCTGC  mCGCAGTTGA  TGAACCTTGs
      1201  CTTTAYCGGC  CCACTrrAAC  rCaSTCGGAC  TTTGCGTTGC  CATCGGTCTG
      1251  GCGCGTGTA  TCAATGCCGG  ATTGTTGTTT  TACCTGTTGC  GCAGACACGG
      1301  TATTTACCAA  CCTGG. CAAG  GGTGGGCAG  CGTTCTT. AG  CAAAATGCT
      1351  GcTCTCGCTC  GCCGTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```

45      1  MNMLGALAKV  GSLTMVSRVL  GFVRDVIAR  AFGAGMATDA  FFVAFKLPNL
      51  LRRVFAEGAF  AQAFVPILAE  YKETRSKEAX  EAFIRHVAGM  LSFVLVIVTA
      101  LGILAAPWVI  YVSAPSEAQD  ADKFQLSIDL  LRITFPYILL  ISLSSFVGSV
      151  LNSYHKFGIP  AFTPXFLNVS  FIVFALFFVP  YFDPVVTAXA  WAVFVGILQ
      201  LXFQLPWLAK  LGFLKLPKLS  FKDAAVNRVM  KOMAPAILGV  SVAQVSLVIN
      251  TIFASYLQSG  SVSWMYADR  MMELPSGVLG  AALGTILLPT  LSKHSANQDT
      301  EQFSALLDWG  LRLCMLTLP  AAVGLAVLSF  PLVATLFMYR  XFTLFDAQMT
      351  QHALIAYSFG  LIGLIMIKVL  APGFYARQNI  XXPVKIAIFT  LICXQLMNLX
      401  FXGPLXXIGL  SLAIGLGACI  NAGLLFYLLR  RHGIYQPXQG  LGSVLXQKCC
      451  SRSP*
  
```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

```

55      1  ATGAATATGC  TGGGAGCTTT  GGCAAAAGTC  GGCAGCCTGA  CGATGGTGTG
      51  GCGCGTTTTG  GGATTTGTGC  GCGATACGGT  CATTGCGCGG  GCATTCGGCG
  
```

-116-

5
10
15
20
25
30

```

101  CGGGTATGGC  GACGGATGCG  TTTTTTGTCT  CGTTCAAAC  T G C C C A A C C T G
151  CTTCGCCGCG  TGTTTGCGGA  GGGGGCGTTT  G C C C A A G C G T  T T G T G C C G A T
201  TTTGGCGGAA  TACAAGGAAA  CGCGTTCAAA  A G A G G C G G C G  G A G G C T T T A
251  TCCGCCATGT  GGC G G G G A T G  C T G T C G T T T G  T A C T G G T T A T  C G T T A C C G C G
301  CTGGGCATAC  TTGCCGCGCC  TTGGGTGATT  T A T G T T T C C G  C A C C C G G T T T
351  TGCCCAAGAT  GCCGACAAAT  TTCAGCTCTC  C A T C G A T T T G  C T G C G G A T T A
401  CGTTTCCTTA  TATATTATTG  ATTTCCCTGT  C T T C A T T T G T  C G G C T C G G T A
451  CTCAATTCTT  ATCATAAGTT  CGGCATTCCG  G C G T T T A C G C  C C A C G T T T C T
501  GAACGTGTCG  TTTATCGTAT  TCGCGCTGTT  T T T C G T G C C G  T A T T T C G A T C
551  CGCCCGTTAC  CGCGCTGGCG  TGGGCGGTCT  T T G T C G C G C G  C A T T T T G C A A
601  CTCGGCTTCC  AACTGCCCTG  GCTGGCGAAA  C T G G G C T T T T  T G A A A C T G C C
651  CAAACTGAGT  TTCAAAGATG  CGGCGGTCAA  C G C G T G A T G  A A A C A G A T G G
701  CGCCTGCGAT  TTTGGGCGTG  AGCGTGCGCG  A G G T T T C T T T  G G T G A T C A A C
751  ACGATTTTCG  CGTCTTATCT  GCAATCGGGC  A G C G T T T C A T  G G A T G T A T T A
801  CGCCGACCGC  ATGATGGAAG  TGCCAGCGGG  C G T G C T G G G G  G C G G C A C T C G
851  GTACGATTTT  GCTGCCGACT  TTGTCCAAAC  A C T C G G C A A A  C C A A G A T A C G
901  GAACAGTTT  CCGCCCTGCT  CGACTGGGGT  T T G C G C T G T  G C A T G C T G C T
951  GACGCTGCCG  GCGGCGGTCT  GACTGGCGGT  G T T G T C G T T C  C C G C T G G T G G
1001  CGACGCTGTT  TAGTACCGC  GAATTTACGC  T G T T T G A C G C  G C A G A T G A C G
1051  CAACACGCGC  TGATTGCGTA  TTCTTTCGGT  T T A A T C G G C T  T A A T C G A T
1101  TAAAGTGTG  GCACCCGGCT  TCTATGCGCG  G C A A A C A T C  A A A A C G C C C G
1151  TCAAAATCGC  CATCTTCACG  CTCATCTGCA  C G C A G T T G A T  G A A C C T T G C C
1201  TTTATCGGCC  CACTGAAACA  CGTCGGACTT  T C G C T T G C C A  T C G G T C T G G G
1251  CGCGTGATC  AATGCCGAT  TGTTGTTT  T G T T G T T T A  C C T G T T G C G C  A G A C A C G G T A
1301  TTTACCAACC  TGGCAAGGGT  TGGGCAGCGT  T C T T A G C A A A  A A T G C T G C T C
1351  TCGCTCGCCG  TGATGTGCGG  CGGACTGTGG  G C A G C G C A G G  C T T A C C T G C C
1401  GTTTGAATGG  GCGCACGCCG  GCGGAATGCG  G A A A G C G G G G  C A G C T C T G C A
1451  TCCTGATTGC  CGTCGGCGGC  GGAATGTATT  T C G C A T C A C T  G C G G G C T T T G
1501  GGCTTCCGTC  CGCGCCATTT  CAAACGCGTG  G A A A A C T G A

```

30 This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

35
40

```

1  MNMLGALAKV  GSLTMVSRVL  GFVRDVIAR  AFGAGMATDA  FFVAFKLPNL
51  LRRVFAEGAF  AQAFVPILAE  YKETRSKEAA  EAFIRHVAGM  LSFVLVIVTA
101  LGILAAPWVI  YVSAPGFAQD  ADKFQLSIDL  LRITFPYILL  ISLSSFVGSV
151  LNSYHKFGIP  AFTPTFLNVS  FIVFALFFVP  YFDPPVTALA  WAFVGGILQ
201  LGFQLPWLAK  LGFLKLPKLS  FKDAAVNRVM  KQMAPAILGV  SVAQVSLVIN
251  TIFASYLQSG  SVSWMYADR  MMELPSGVLG  AALGTILLPT  LSKHSANQDT
301  EQFSALLDWG  LRLCMLLTLP  AAVGLAVLSF  PLVATLFMYR  EFTLFDAQMT
351  QHALIAYSFG  LIGLIMIKVL  APGFYARONI  KTPVKIAIFT  LICTQLMNL
401  FIGPLKHVGL  SLAIGLGACI  NAGLLFYLLR  RHGIYQPGKG  WAAFLAKMLI
451  SLAVMCGGLW  AAQAYLPFEW  AHAGGMRKAG  QLCILIAVGG  GLYFASLAAL
501  GFRPRHFKRV  EN*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

45
50
55
60

```

Orf20 1  MNMLGALAKV  GSLTMVSRVL  GFVRDVIAR  AFGAGMATDA  FFFVAFKLPNLLRRVFAEGAF 60
MviN 14  MNLLKSLAAVSSMTMFSRVLGFAIDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73

Orf20 61  AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPSFAQD 120
MviN 74  SQAFVPILAEYKSKQGEETATRFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133

Orf20 121  ADKFQLSIDLLRITFPYILLISLSSFVGSVLSYHKFGIPAFPTXFLNVSFIVFALFFVP 180
MviN 134  ADKFALTQQLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193

Orf20 181  YFDPPVTAXAWAVFVGGILQLXFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV 240
MviN 194  YFNPPVLALAWAVTVGGVLQVLVYQLPYLKKIGMLVLPINFRDGTAMRVVVKQMGPAILGV 253

Orf20 241  SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT 300
MviN 254  SVQISLIINTIFASFLASGSVSWMYADRLMEFPGVLGVALGTILLPSLSKSFASGNH 313

```

-117-

Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 360
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G
 MviN 314 DEYCRIMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGFKAFAFDAAMTQRALIAYSVG 373

5 Orf20 361 LIGLIMIKVLAPGFYARONIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXXCI 420
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+
 MviN 374 LIGLIVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Orf20 421 NAGLLFYLLRRHGIYQXPQG 440
 NA LL++ LR+ I+ P G
 MviN 434 NASLLYWQLRKQNIPTQPG 453

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N.*

15 *meningitidis*:

	10	20	30	40	50	60
orf20.pep	MNMLGALAKVGS	LTMSRVLG	FVRD	TVIARAF	GAGMATDA	FFVAFKLPNLLRRVFAEGAF
orf20a	MNMLGALAKVGS	LTMSRVLG	FVRD	TVIARAF	GAGMATDA	FFVAFKLPNLLRRVFAEGAF
	70	80	90	100	110	120
orf20.pep	AQAFVPILAEYK	ETRSKEAXEAF	IRHVAGML	SFVLVIVT	ALGILAAP	PWVIYVSAPSFAQD
orf20a	AQAFVPILAEYK	ETRSKEATEAF	IRHVAGML	SFVLVIVT	ALGILAAP	PWVIYVSAPGFAKD
	130	140	150	160	170	180
orf20.pep	ADKFQLSIDLLR	ITFPYILLIS	LSFVGS	VLNSYHK	FGIPAF	TPXFLNVSFIVFALFFVP
orf20a	ADKFQLSIDLLR	ITFPYILLIS	LSFVGS	VLNSYHK	FSIPAF	TPXFLNVSFIVFALFFVP
	190	200	210	220	230	240
orf20.pep	YFDPPVTAXAW	AVFVGGILQ	LQFQLP	WLAKLGL	FLKLPKLS	FKDAAVNRVMKQMAPAILGV
orf20a	YFDPPVTALAW	AVFVGGILQ	LQFQLP	WLAKLGL	FLKLPKLS	FKDAAVNRVMKQMAPAILGV
	250	260	270	280	290	300
orf20.pep	SVAQVSLVINTI	FASYLQSG	SVSWMY	YADRMEL	PSGVLGA	ALGTILLPTLSKHSANQDT
orf20a	SVAQISLVINTI	FASYLQSG	SVSWMY	YADRMEL	PGGVLGA	ALGTILLPTLSKHSANQDT
	310	320	330	340	350	360
orf20.pep	EQFSALLDWGL	RLCMLLTLP	AAVGLAVL	SFPLVATL	LFMYRXFT	LFDAQMTQHAIAYSFG
orf20a	EQFSALLDWGL	RXCMLLTLP	AAVGMVAVL	SFPLVATL	LFMYREF	LFDAQMTQHAIAYSFG
	370	380	390	400	410	420
orf20.pep	LIGLIMIKVLAP	GFYARONIX	XPVKIAI	FTLICXQ	LMNLXFX	GPLXXIGLSLAIGLGACI
orf20a	LIGLIMIKVLAP	GFYARONIK	TPVKIAI	FTLICTQ	LMNLAFI	GPLKHVGLSLAIGLGACI
	430	440	450	460	470	480
orf20.pep	NAGLLFYLLRR	HGIYQXPQ	GLGSVLX	QKCCSR	SPX	
orf20a	NAGLLFYLLRR	HGIYQPGK	GWAFLAK	MLLSLA	VMGGGLY	AAQIWL

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

65 1 ATGAATATGC TGGGAGCTTT GGTAAAAGTC GGCAGCCTGA CGATGGTGTG
 51 GCGCGTTTGTG GGATTTGTGC GCGATACGGT CATTGCGCGC GCATTGCGCG
 101 CAGGCATGGC GACGGATGCG TTCTTTGTGC CGTTCAAACCT GCCCAACCTG

-118-

5
10
15
20
25

```

151 CTTGCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGACG GAGGCTTTTA
251 TCCGCCATGT GCGGGGGATG CTGTGCTTTG TACTGGTCAT CGTTACCGCG
301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
351 TGCCAAAGAT GCCGACAAAT TTCAGCTCTC TATCGATTG CTGCGGATTA
401 CGTTTCCTTA TATCTTATTG ATTTCACTTT CCTCTTTTGT CGGCTCGGTA
451 CTCAATTCCT ATCATAAATT CAGCATTCCT GCGTTTACGC CCACGTTCCCT
501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCCG TATTTTCGATC
551 CTCCCGTTAC CGCGCTGGCT TGGGCGGTTT TTGTCGGCGG CATTTTGCAA
601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGTTTTT TGAAACTGCC
651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
701 CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGATTTCTTT GGTGATCAAC
751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATTCATTA
801 CGCCGACCGC ATGATGGAAC TGCCCGGGCG CGTGCTGGGG GCGGCACTCG
851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCNTGT GCATGCTGCT
951 GACGCTGCCG GCGGCGGTCG GAATGGCGGT GTTGTCTGTT CCGCTGGTGG
1001 CAACCTTGTT TATGTACCGA GAATTCACGC TGTGTGACGC GCAGATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATCATGAT
1101 TAAAGTGTTG GCGCCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
1151 TCAAAATCGC CATCTTCACG CTCATTTGCA CGCAGTTGAT GAACCTTGCC
1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
1251 CGCGTGATC AATGCCGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
1301 TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTGGCAA AATGCTGCTC
1351 TCGTCGCGCG TGATGGGAGG CGGCCTGTAT GCCGCCAAA TCTGGCTGCC
1401 GTTCGACTGG GCACACGCCG GCGGAATGCA AAAGGCCGCC CGGCTCTTCA
1451 TCCTGATTGC CGTCGGCGGC GGAATGTATT TCGCATCACT GGCGGCTTTG
1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 118>:

30
35
40

```

1 MNMLGALVKV GS LTMVSRVL GFVRDVIAR AFGAGMATDA FVFAFKLPNL
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLVIVTA
101 LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151 LNSYHKFSIP AFTPTFLNVS FIVEALFFVP YFDPVPTALA WAVFVGGILO
201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN
351 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
401 EQFSALLDWG LRXCMLLTLF AAVGMAVLSF PLVATLFMYR EFTLFDAQMT
451 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
501 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
551 SLAVMGGGLY AAQIWLPPDW AHAGGMQKAA RLFILIAVGG GLYFASLAAL
501 GFRPRHFKRV ES*

```

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

45
50
55
60
65

```

          10      20      30      40      50      60
orf20a.pep MNMLGALVKVGS LTMVSRVLGFVRDVIARA FGAAGMATDAFFVFAFKLPNLLRRVFAEGAF
          10      20      30      40      50      60
orf20-1     MNMLGALAKVGS LTMVSRVLGFVRDVIARA FGAAGMATDAFFVFAFKLPNLLRRVFAEGAF
          70      80      90      100     110     120
orf20a.pep  AQAFVPILAEYKETRSKEATEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFAKD
          70      80      90      100     110     120
orf20-1     AQAFVPILAEYKETRSKEAAEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFAQD
          130     140     150     160     170     180
orf20a.pep  ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIP AFTPTFLNVSFIVFALFFVP
          130     140     150     160     170     180
orf20-1     ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIP AFTPTFLNVSFIVFALFFVP
          190     200     210     220     230     240
orf20a.pep  YFDPVPTALAWAVFVGGILQLGFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV
          190     200     210     220     230     240
orf20-1     YFDPVPTALAWAVFVGGILQLGFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV
          250     260     270     280     290     300
orf20a.pep  SVAQISLVINTIFASYLQSGSVSWMYADRM MELPGGVLG AALGTILLPTLSKHSANQDT

```

10

15

20

25

30 ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N. gonorrhoeae*:

35

40

45

50

55

60

BNSDOCID: <WO 9924578A2_1_>

-120-

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGSI
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPVPTALA WAVFVGILQ
 201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KOMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGQG LGQPSWRKCC
 451 SRSP*

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
 101 CCGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACT GCCCAACCTG
 151 CTTGCGCGCG TGTTCGCGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGAag gAGGCTTTTA
 251 TCCGCCACGt tgcgggAatg CTGTCGTTTG TGCTGATcgt cGttacCGCG
 301 CTGGGCATAC TTGCGCGgcc tTGGGTGATT TATGTTtccg CgcccGGCTT
 351 TACCAAGAC TCGGACAAGT TCCAACCTTC CATCAGCCTG CTGCGGATTA
 401 CGTTTCCTTA TATATTATTG ATTTCTTTGT CTTCTTTTGT CGGCTCGATA
 451 CTCAATTCCT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT
 501 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCTGTCCG TATTTTCGATC
 551 CGCCGTTTAC GCGCTGGCG TGGGCGGTTT TTGTGCGCGG TATTTTGCAG
 601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
 651 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG
 701 CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta
 801 cgCCGACCGC ATGATGGAGc tgcgcccGGG CGTGCTGGG GCTGCACTCG
 851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
 951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTCGTTT CCGCTGGTGG
 1001 CGACGCTGTT TATGTACCGA GAATTCACGC TGTTCGACG ACAAATGACG
 1051 CAACACGCGC TGATTCGCTA TTCTTTCGGT TTAATCGGTT TAATTATGAT
 1101 TAAAGTGTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC
 1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCA TCGCCTGGG
 1251 CGCGTGCATC AACGCCGGAT TGTGTCTCTT CCTGTTGCGC AAACACGGTA
 1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGCGGAA AATGCTGCTC
 1351 GCGCTCGCCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCTGCC
 1401 GTTCGAATGG GCGCAGCCG GCGGAATGCG GAAAGCGGG CAGCTCTGCA
 1451 TCCTGATTGC CGTCGGCGGC GGAAGTATT TCGCATCTCT GCGGCTTTG
 1501 GGCTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA

This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGSI
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPVPTALA WAVFVGILQ
 201 LGFLPWLAK LGFLKLPKLN FKDAAVNRVM KOMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAAFLAKMLL
 451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
 501 GFRPRHFKRV ES*

ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

		10	20	30	40	50	60
orf20-1.pep		MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
orf20ng-1		MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
		10	20	30	40	50	60
		70	80	90	100	110	120
orf20-1.pep		AQAFVPILAE	YKETRSKEA	EAFIRHVAGM	LSFVLIVVTA	LGILAAPWVI	YVSAPGFAQD
orf20ng-1		AQAFVPILAE	YKETRSKEA	EAFIRHVAGM	LSFVLIVVTA	LGILAAPWVI	YVSAPGFTKD

-121-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf20-1.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFPTPTFLNVSFIVFALFFVP					
	orf20ng-1	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTPTFLNISFIVFALFFVP					
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf20-1.pep	YFDPPTALAWAVFVGGILQLGFQLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
	orf20ng-1	YFDPPTALAWAVFVGGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV					
		190	200	210	220	230	240
		250	260	270	280	290	300
15	orf20-1.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
	orf20ng-1	SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
		310	320	330	340	350	360
20	orf20-1.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
25		310	320	330	340	350	360
		370	380	390	400	410	420
30	orf20-1.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
		370	380	390	400	410	420
		430	440	450	460	470	480
35	orf20-1.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
	orf20ng-1	NAGLLFFLLRKHGIYRPGRGWAAFLAKMLLALAVMCGGLWAAQACLPEWAHAGGMRKAG					
		430	440	450	460	470	480
		490	500	510			
40	orf20-1.pep	QLCIIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCIIAVGGGLYFASLAALGFRPRHFKRVESX					
		490	500	510			

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S. typhimurium*:

45	sp P37169 MVIN_SALTY_VIRULENCE_FACTOR_MVIN_pir S40271 mviN protein - Salmonella typhimurium gi 438252 (Z26133) mviB gene product [Salmonella typhimurium] gnl PID d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220 Identities = 309/467 (66%), Positives = 368/467 (78%)
50	Query: 1 MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF Sbjct: 14 MNLLKSLAAVSSMTMF SRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
55	Query: 61 AQAFVPILA EYK ETRSKEATEAFIRHVAGMLS FVLIVVTALGILAAPWVIYVSAPGFTKD 120 +QAFVPILA EYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF Sbjct: 74 SQAFVPILA EYKSKQGE EATRIFVAYVSGLLT LALAVVTVAGMLAAPWVIMVTAPGFADT 133
60	Query: 121 ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTPTFLNISFIVFALFFVP 180 ADKF L+ LLRITFPYILLISL+S VG+ILN++++F IPAF PTFLNIS I FALF P Sbjct: 134 ADKFALT TQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISIMIGFALFAAP 193
65	Query: 181 YFDPPTALAWAVFVGGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV 240 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAIILGV Sbjct: 194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRI NFRDTGAMRVVKQMGPAIILGV 253
70	Query: 241 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT 300 SV+QISL+INTIFAS+L SGSVSWMYADR+ME GVLG ALGTILLP+LSK A+ + Sbjct: 254 SVSQISLIINTIFASFLASGSVSWMYADRLMEFP SGVLGVALGTILLPSLSKSFASGNH 313

-122-

Query: 301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG 360
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G
 Sbjct: 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSFG 373

5 Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420
 LIGLI++KVL A GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+
 Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Query: 421 NAGLLFFLLRKHGIYRPGRWXXXXXXXXXXXXXVMCGGLWAAQACLP 467
 NA LL++ LRK I+ P GW VM L+ +P
 Sbjct: 434 NASLLYWLRLKQNIPTPQPGWMWFLMRLLISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 14/41 (34%), Positives = 23/41 (56%)

15 Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509
 EW+ + + +L ++ G YFA+LA LGF+ + F R
 Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521

- 20 Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

25 1 atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
 201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTAC CGTGGCGAAA
 30 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA TGGACACCAA TCCG..

- 35 This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEXNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNP..

- 40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
 45 201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTAC CGTGGCGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 50 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCAATTAT
 501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT
 551 TGACCGAACG CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
 651 TGCCGGTTTG AGTGGCACGC ACATTCATTT CATCGAGCCG GTCGGCGCGA
 701 ATAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
 801 TTCTCAAGTC AACAAACCGC GCCTCTTGGC TACCGTTTTG GGTGCGAAG
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
 901 TCCGTTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT

-123-

5
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTCCG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
 1051 ACAACCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC
 1101 CGTCAACGGC GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCGATACCG ACAGCGCGCA GGCATTGGGT TGCTTGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

10
 1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVITIGR
 15
 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

20
 1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
 201 GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGCGGAAA
 25
 251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACTTAA GCGGCGANGA
 351 ANTNNGNNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
 401 GTCCGTTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCGT TGGTTGTGAT
 30
 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
 551 TGACCGAGCG TAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
 651 GGCCGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGTTGCAA
 701 AAAAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
 35
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
 801 TTCTCAAGTC AAAAAACCCG GCCTCTTGGC TACCGTTTGG GGTGCGAAAG
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
 901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 40
 1001 AGCTGTTCCG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
 1051 ACGACCCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGT GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCGATACCG ACAGCGCGCA AGCATTGGGT TGCTTGAAT TGGACGAAGA
 45
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
 1301 CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

50
 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PXMKVKEGDA
 51 VKKGQVLFED KXPGVVFTA PVSGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGXEXXX NLIQSGLWTA LRXPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPVVVIKEA XXDFRRXXLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDADNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV
 55
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

60
 orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
 orf22a MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED

-124-

		10	20	30	40	50	60
5	orf22.pep	70	80	90	100	110	120
		KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIAVEKNDEIEFERYAPEALANLSGEEVRR					
	orf22a	70	80	90	100	110	120
		KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYAPEALANLSGXEXXX					
10	orf22.pep	130	140	150			
		NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP					
	orf22a	130	140	150	160	170	180
		NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV					
15	orf22a.pep	10	20	30	40	50	60
		MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED					
	orf22-1	10	20	30	40	50	60
		MIKIKKGLNLPIAGRPEQVYDGPVITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED					
20	orf22a.pep	70	80	90	100	110	120
		KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYAPEALANLSGXEXXX					
	orf22-1	70	80	90	100	110	120
		KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYAPEALANLSGEEVRR					
25	orf22a.pep	130	140	150	160	170	180
		NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV					
	orf22-1	130	140	150	160	170	180
		NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVVIKEAAEDFKRGLLV					
30	orf22a.pep	190	200	210	220	230	240
		LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTVWTI					
	orf22-1	190	200	210	220	230	240
		LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTVWTI					
35	orf22a.pep	250	260	270	280	290	300
		NYQDVIAIGRLFATGRLNTERVIALGGSQVNKPRLRLTVLGAKVSQITAGELVDADNRVI					
	orf22-1	250	260	270	280	290	300
		NYQDVITIGRLFATGRLNTERVIALGGSQVNKPRLRLTVLGAKVSQITAGELVDTDNRVI					
40	orf22a.pep	310	320	330	340	350	360
		SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK					
	orf22-1	310	320	330	340	350	360
		SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK					
45	orf22a.pep	370	380	390	400	410	420
		LFKFTTAVNGGDRAMVPIGTYERVMPDILPTLLRLDLIVGDTDSAQALGCLELDEEDLA					
	orf22-1	370	380	390	400	410	420
		LFKFTTAVNGGDRAMVPIGTYERVMPDILPTLLRLDLIVGDTDSAQALGCLELDEEDLA					
50	orf22a.pep	430	440				
		LCSFVCPGKYEXGPLLRLKVLETXEKEGX					
	orf22-1	430	440				
		LCSFVCPGKYEYGPLLRLKVLETIEKEGX					

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

65 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYVGM R PSMKIKEGEA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYVPEAL AKLSSEKVR. NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF

-125-

151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFVTGRLNTE RVVALGGLQV NKPRLRLTVL GAKVSQLTAG ELVDADNRVI
 301 SGSVLNGAIA QGAHDYLGRI HN*

5 Further work identified complete gonococcal gene <SEQ ID 131>:

1 ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTAGT
 10 201 ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTAC CGTGGCGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATTGA GCAGCGAAAA
 351 AGTGCGCCGC AACCTGATTC AATCAGGCTT ATGGACTGCG CTTCGCACCC
 401 GTCCGTTTCA CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC
 15 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTATCAT
 501 CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC
 551 TGACCGAAGC TAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC
 651 TGCCGGCTTG AGTGCGACGC ACATTCATTT CATCGAGCCA GTCGGCGCGA
 20 701 ATAAAACCGT GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT
 751 TTGTTTCGTAA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CCTTGGGCGG
 801 CCTGCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAGG
 851 TGTCTCAACT TACCGCCGCG GAATTTGGTTG ACGCGGACAA CCGCGTGATT
 901 TCCGGTTCGG TATTGAACGG TGCGATTGCA CAAGGCGCGC ATGATTATTT
 25 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTTCG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGC
 1051 ACCACTCTCG GCCATTTCTT AAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGC GCGCAGCCGC CCATGGTACC GATCGGCACT TATGAGCGCG
 1151 TAATGCCGTT GGACATCCTG CCTACCTTGC TTTTGCGCGA TTTAATCGTC
 30 1201 GCGGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

1 MIKIKKGLNL PIAGRPEQVI YDGPATEIVA LLGEEYVGM RPSMKIKEGEA
 35 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYVPEAL AKLSSEKVR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFVTGRLNTE RVVALGGLQV NKPRLRLTVL GAKVSQLTAG ELVDADNRVI
 40 301 SGSVLNGAIA QGAHDYLGRI HNQISVIEEG RSKELEFGWA POPDKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLKVL ETIEKEG*

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa

45 overlap with ORF22ng:

orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPATEIVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60
 orf22ng MIKIKKGLNLPIAGRPEQVIYDGPATEIVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 60
 50 orf22.pep KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR 120
 orf22ng KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 120
 orf22.pep NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158
 55 orf22ng NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 180

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

60 orf22-1.pep MIKIKKGLNLPIAGRPEQAVYDGPATEIVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED

10

15

20

25

30

35

40

45

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

50

55

60

gi|1185395 (U24492) 48 kDa outer membrane protein (Actinobacillus pleuropneumoniae)
Length = 449

64

BNSDOCID: <WO 9924578A2_1_>

-127-

Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

Query: 1 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMVKVEGDAVKKGQVLFED 60
 5 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 61 KKPXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXX 120
 10 Sbjct: 61 KKNPGVVFTAPASGT VVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180
 Sbjct: 121 NLI+SGLWTA R RPFSSK+PA+DA P +IFVNAMDTNPLAADP VV+KE DF+ V 180

Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTV 237
 15 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 238 WTINYQDVIAIGRLFATGRINTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADN 297
 20 Sbjct: 241 W+NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQ+TA EL +N 300

Query: 298 RVISGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFL 357
 25 Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KNKLFKFTTAVNGGDRAMVPIGT YERVMXXXXXXXXXXXXXXXXXVGD TDSAQXXXXXXXXXX 417
 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 418 XXXXXSFVCPGKYEXGPLL RKVLETXEKEG 447
 30 ++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A. pleuropneumoniae*:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus
 35 pleuropneumoniae] Length = 449
 Score = 555 bits (1414), Expect = e-157
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

Query: 27 MIKIKKGLNLPIAGRPEQVIYDGP AITEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 86
 40 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 146
 45 Sbjct: 61 KKNPGVVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++ 120

Query: 147 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTV IIEKAAEDFKRGLLV 206
 50 Sbjct: 121 NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNPLAADP V+KE DFK GL V 180

Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTV 263
 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTGRINTERVVALGGLQVKNPRLRLTVLGAKVSQITAGELVDADN 323
 55 Sbjct: 241 W+NYQDVIAIG+LF TG L T+R+++L G QV PRL+RT LGA +SQLTA EL +N 300

Query: 324 RVISGSVLNGAIAQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFL 383
 60 Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KNKLFKFTTAVNGGDRAMVPIGT YERVMXXXXXXXXXXXXXXXXXVGD TDSAQXXXXXXXXXX 443
 65 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 444 XXXXXSFVCPGKYEYGPLLRKVLETIEKEG 473
 ++VCPGK YGP+LR LE IEKEG

Sbjct: 420 DLALCTYVCPGKNNYGPMLRAALEKIEKEG 449

70

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```

1   .GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
101 TTA CTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
151 TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
251 TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20 401 GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGGsGC TTTmTTGsw CAkCATCTTT TTTGCCGCAC AGTTTGTCGC
501 ATTTTTTAAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551 CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTGTT TATCGTTTTT
601 ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
25 651 ATGGGCGGTA ACTGCCCGCA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701 ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751 AATATTATTA CGCCGATGAT GAGTTATTTT GGGCTGATTA TGGCGACGGT
801 GrkCmmTAC AAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851 TGCCGTATTC CGCTTTCTTC TTGATTGCGt GGATTGCCTT ATTCTGCATT
30 901 TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
951 TCCCGCACCT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

1   .AXXIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
35 101 TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151 TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIGF
201 ILICAFINLM IGSASQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
251 NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
301 WVFVLGLPVG PGAPTYPAP *

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

1   ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTTT ATTATTTTCA
101 TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
151 GTCCCCGATC CGCGCCCTGT TGGTGGGAAA GGACGTGCCG ATGACGGTTT
45 201 GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTATC AAAATCTGA
251 CGCATACCGT TAAAAATTTT ACCGGTTTCG CGCCGTTGGG AACGGTGTG
301 GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
351 ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAACCTC ACTACTTTTA
401 TGGTTGTTTT TACAGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
50 451 GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
501 TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTTAT

```

5
10
15
20

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

25

30

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N. meningitidis*:

40
45
50
55
60

-130-

```

orfl2a      IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY
              420      430      440      450      460      470

5    orfl2.pep  KKDAGVGT LIXMMLPYSAFFLI AWIALFCIWV FVLGLPVGPGAPT FYPAPX
              280      290      300      310      320
orfl2a      KKDAGVGT LISMMLPYSAFFLI AWIALFCIWV FVLGLPVGPGAPT FYPAPX
              480      490      500      510      520

```

The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

```

10      1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTTCA
     101  TTGTGTTATT GCTGATTGCC TCTGCCCGCG GTGCGTATTT CGGACATATCC
     151  GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGTTTT
     201  GATTACAGTT GTCAGCCTGC TCGATGCTGA CGGTTTGATC AAAATCCTGA
     251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCGGTGGT AACGGTGTGT
     301  GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAAATCTCC ACGCAAATC ACTACTTTTA
     401  TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
     451  TTCGTCTTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCGGCCA
     501  TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTTATT
     551  CCGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTGGC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTATT TGAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
     751  GATTTGTGAC AAGAAGAAA AGACATTCTGA CATTCCAATG AAATCACGCC
     801  TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTGT GTTGCCTTAT
     851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTTC CGGTTCGCGC TTTTAAAAAT CAATTGTTGT
     951  TTTTATTTTC TTGTTGTTG CACTGCCGGG CATTGTTTAT GGCCGGGTAA
    1001  CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
    1051  ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTGTG CCGCACAGTT
    1101  TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
    1151  GGGCGACGTT CTTAAAAGAA GTCGGCTTGG GCGGCACGGT GTTGTTTATC
    1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCCTC
    1251  CGCGCAATGG GCGGLAACTG CGCCGATTTT CGTCCCTATG CTGATGTTGG
    1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CCGTGATTCC
    1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCCGGC TGATTATGGC
    1401  GACGGTGATC AAATACAAA AAGATGCGGG CGTGGGTACG CTGATTCTTA
    1451  TGATGTTGCC GTATTCGCTT TTCTTCTTGA TTGCGTGGAT TGCCTTATTC
    1501  TGCATTTGGG TATTTGTTTT GGGCCTGCCC GTCGGTCCCG GCGCGCCAC
    1551  ATTCTATCCC GCACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 138>:

```

45      1  MSQTD TQRDGR FLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAAGAYFGLS
      51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
     101  VSLLGVGIAE KSGLISALMR LLTKSPRKL TTFMVVFTGI LSNTASELGY
     151  VVLIPLSAII FHSLGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
     201  QQAAQIIHPD YVVGPEANWF FMVASTEVIA LIGYFVTEKI VEPQLGPYQS
     251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
     301  PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
     351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLEFI
     401  GFILICAFIN LMIGSASQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
     451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
     501  CIWVFLGLP VGPAGPTFYP AP*

```

55 ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

```

60      10      20      30      40      50      60
orfl2a.pep  MSQTD TQRDGR FLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAAGAYFGLS VPDPRPVGAK
              |||
orfl2-1      MSQTD TQRDGR FLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS VPDPRPVGAK
              10      20      30      40      50      60

              70      80      90      100     110     120
orfl2a.pep  GRADDGLIHV VSLLDADGLI KILTHTVKNFTGFAPLGTVL VSLLGVGIAE KSGLISALMR
              |||
orfl2-1      GRADDGLIYV VSLLDADGFI KILTHTVKNFTGFAPLGTVL VSLLGVGIAE KSGLISALMR

```


-131-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf12a.pep	LLLT	KSPR	KLTF	FMVV	FTGIL	SNTASE
	orf12-1	LLLT	KSPR	KLTF	FMVV	FTGIL	SNTASE
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf12a.pep	GGYS	ANLFL	GTID	PLLAG	ITQQA	AI
	orf12-1	GGYS	ANLFL	GTID	PLLAG	ITQQA	AI
		190	200	210	220	230	240
		250	260	270	280	290	300
15	orf12a.pep	VEPQ	LGPY	QSDLS	QEEK	DIRHS	NEIT
	orf12-1	VEPQ	LGPY	QSDLS	QEEK	DIRHS	NEIT
		250	260	270	280	290	300
		310	320	330	340	350	360
20	orf12a.pep	PETG	LVS	GS	PFLKS	IVVF	IFLL
	orf12-1	PETG	LVS	GS	PFLKS	IVVF	IFLL
		310	320	330	340	350	360
		370	380	390	400	410	420
25	orf12a.pep	IFFA	QFVA	FFNWT	NIGQY	IAVK	GATFL
	orf12-1	IFFA	QFVA	FFNWT	NIGQY	IAVK	GATFL
		370	380	390	400	410	420
		430	440	450	460	470	480
30	orf12a.pep	AVTA	PIFV	PMLML	LAGYA	PEVIQ	AAYR
	orf12-1	AVTA	PIFV	PMLML	LAGYA	PEVIQ	AAYR
		430	440	450	460	470	480
		490	500	510	520		
35	orf12a.pep	LISM	LPYS	AFFLI	AWIAL	FCI	WV
	orf12-1	LISM	LPYS	AFFLI	AWIAL	FCI	WV
		490	500	510	520		

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N. gonorrhoeae*:

	orf12.pep	AXXI	IHPX	XVVG	PEAN	WFFM	VAST	FVIALI	30
50	orf12ng	AAAF	AGVS	GGYS	ANLFL	GTID	PLLAG	ITQQA	232
	orf12.pep	GYFV	TEKIV	EPQL	GPYQ	SDLS	QEEK	DIRHS	90
	orf12ng	GYFV	TEKIV	EPQL	GPYQ	SDLS	QEEK	DIRHS	292
55	orf12.pep	PADG	ILRHP	PETG	LVS	GS	PFLKS	IVVF	150
	orf12ng	PADG	ILRHP	PETG	LVS	GS	PFLKS	IVVF	352
60	orf12.pep	TLXL	XLXX	IFFA	QFVA	FFNWT	NIGQY	IAVK	210
	orf12ng	TLGL	YLVI	IFFA	QFVA	FFNWT	NIGQY	IAVK	412
65	orf12.pep	IGS	ASAQ	WAVT	APIF	VPML	MLAG	YAPE	270
	orf12ng	IGS	ASAQ	WAVT	APIF	VPML	MLAG	YAPE	472

```

orfl2.pep      KKDAGVGTLIXMMLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFFYPAP  320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orfl2ng        KKDAGVGTLISMMLPYSAFFLIWIALFCIWVFLGLPVGPGTPTFFYPVP  522

```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTTCA
     101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATTT CGGACTATCC
     151  GTCCCCGATC CGCGTCCGTG TGGGGCGAAA GGACGTCCCG ATGACGGTTT
     201  GATTACAGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCCTGA
     251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTG
     301  GTTCTTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAAATCCCC ACGCAAATC ACTACTTTTA
     401  TGGTTGTTTT TACAGGGATT TTATCCAATA CGGCTTCTGA ATTTGGCTAT
     451  GTCGTCTCTA TCCCTTTGTC CGCCGTCATC TTTCATTTCG TCGGCCGCCA
     501  TCCGCTTGCC GGTTTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTATT
     551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTTAC TGAAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
     751  GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
     801  TTTGGAATAT AAAGGATTAA TTTGGGCAAG CGTGGTGTG GTTGCTTAT
     851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTGC CGGTTTCGCG TTTTAAATAT CGATTGTTG
     951  TTTTATTTTC TTGTTGTTTG CGCTGCCGGG CATTGTTTAT GGCCGATAA
    1001  CCCGAAGTTT GCGCGCGCAA CGGGAAGTCG TTAATCGCAT GGCCGAATCG
    1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
    1101  TGTGCGATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAG
    1151  GGGCGGTGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTATTATC
    1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
    1251  CGCGCAATGG GCGGTAAC TGCCGATTTT CGTCCCTATG CTGATGTTGG
    1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CCGTGATTCC
    1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCCGGC TGATTATGGC
    1401  GACGGTAATC AAATACAAAA AAGATGCGGG CGTAGGCACG CTGATTTCTA
    1451  TGATGTTGCC GTATTCGCGT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
    1501  TGCATTTGGG TATTTGTTTT GGGTCTGCCC GTCGGTCCCG GCACACCCAC
    1551  ATTCTATCCG GTGCCCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

      1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
     51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
    101  VSLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LNSTASELGY
     151  VVLIPLSAVI FHSLSGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
    201  QQAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
     251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
     301  PETGLVAGSP FLKSIIVFIF LFLALPGIVY GRITRSLRGE REVVNAMAE
    351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGAVFLKK FRLGGSVLFI
     401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPOV IQAAYRIGDS
     451  VTNIITPMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
     501  CIWVFLGLP VPGTPTFFYP VP*

```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50      10      20      30      40      50      60
orfl2-1.pep MSQTDTRDGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
orfl2ng      MSQTDARRSGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK

55      10      20      30      40      50      60
orfl2-1.pep GRADDGLIIVSLLNADGF I KILTHTVKNFTGFAPLGTVLVSLGVGIAEKSLGISALMR
orfl2ng      GRADDGLIHVSLLDADGLI KILTHTVKNFTGFAPLGTVLVSLGVGIAEKSLGISALMR

60      70      80      90      100     110     120
orfl2-1.pep LLLTKSPRKLTTFMVFTGILSNTASELGYVVLIPLSAII FHSLSGRHPLAGLAAAFAGVS
orfl2ng      LLLTKSPRKLTTFMVFTGILSNTASELGYVVLIPLSAII FHSLSGRHPLAGLAAAFAGVS

65      130     140     150     160     170     180
orfl2-1.pep LLLTKSPRKLTTFMVFTGILSNTASELGYVVLIPLSAII FHSLSGRHPLAGLAAAFAGVS
orfl2ng      LLLTKSPRKLTTFMVFTGILSNTASELGYVVLIPLSAII FHSLSGRHPLAGLAAAFAGVS

```

-133-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf12-1.pep	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
	orf12ng	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKI					
		190	200	210	220	230	240
10	orf12-1.pep	250	260	270	280	290	300
	orf12ng	VEPQLGPGYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
		250	260	270	280	290	300
15	orf12-1.pep	310	320	330	340	350	360
	orf12ng	PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI					
		310	320	330	340	350	360
20	orf12-1.pep	370	380	390	400	410	420
	orf12ng	IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQW					
		370	380	390	400	410	420
25	orf12-1.pep	430	440	450	460	470	480
	orf12ng	AVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
		430	440	450	460	470	480
30	orf12-1.pep	490	500	510	520		
	orf12ng	LISMMLPYSAFFLIAWIALFCIWWFVLGLPVGPGAPTFFYPAPX					
		490	500	510	520		

In addition, ORF12ng shows significant homology with a hypothetical protein from *E.coli*:

40	sp P46133 YDAH_ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION
	>gi 1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli]
	Length = 510
	Score = 329 bits (835), Expect = 2e-89
	Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)
45	Query: 8 RSGRFLRTVEWLGNNLPHPVVXXXXXXXXXXASAVGAYFGLSVPDPRPVGAKGRADDGL 67
	+SG+ VE +GN +PHP +A+ +FG+S +P D
	Sbjct: 13 QSGKLYGWVERIGNKVPHFLLFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64
50	Query: 68 IHVVSLLDADGLIKILHTVKNFTGFAPXXXXXXXXXXIAEKSGLISALMRLLLT KSP 127
	+ V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + +
	Sbjct: 65 VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124
55	Query: 128 RKLTTFMVFTGILSNTASELGYYVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187
	+ ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL
	Sbjct: 125 ARYASYMVLFIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVCGFTANL 184
60	Query: 188 FLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247
	+ T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG
	Sbjct: 185 LIVTTDVLLSGISTEAAAANFPMHVSVIDNWFYMASSVVVLTIVGGLITDKIIEPRLGQ 244
65	Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRHPETGLVA 307
	+Q + ++ + + S GL AGVV + A +A ++P +GILR P V
	Sbjct: 245 WQGNSDEKLQTLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298
70	Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLXXXXXXXXXX 367
	.SPF+K IV I L F + + YG TR++R + ++ + M E M + ++
	Sbjct: 299 PSPFIKGIVPLIILFFFVVSLEYGIATRTIRRQADLPHLMIEPMKEMAGFIVMVFPPLAQF 358
	Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQWAVTAPIF 427
	NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF

-134-

Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418
 Query: 428 VPMLMLAGYAPEVIQAAAYRIGDSVTNIITPMSYFGLIMATVIKYKKDAGVGTLSMMLP 487
 VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP
 Sbjct: 419 VPMFMLLGfHPAFAQILFRIADSSVLPLAPVSPFVPLFLGFLQRYKPDALGTYYSVLVP 478
 Query: 488 YSAFFLIAWIALFCIWVFLGLPVGPG 514
 Y FL+ W+ + W +++GLP+GPG
 Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 17

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

```

1  ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA
51  GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAAACA GGTTTTTTTC
101 ATGGCATTTC GGTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
151 ATgGCTTCGC GCAGTGCCTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
201 GACGGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTgTT
251 CTTCAGACGG CAGCAGGTCG GTTTGTGTGT ACACCTTgAT GCACGGAAaTA
301 TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
351 TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
25  401 gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCGGC
451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTGCGG
501 ACT..

```

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

```

1  ..TAGAAGXXVF VFVTD SQVEV FGNIQTAVET GFFHGISVSS VFGAAQDSA
30  51  MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRs VLLYTLMHGI
101 SPAWISCSTF STSSICPLF GAAASTTCSS TSACAVSSSV AEKAEISLCG
151 RXLTNP TVSV RIMLHSG..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N.meningitidis*:

```

                                10      20      30
orfl4.pep                      TAGAAGXXVFVFVTD SQVEVFGNIQTAVET
                                |:| | | | | | | | | |:| | | | | | | | |
40  orfl4a                      GRQLGFLRVGGALFVITAQARVNNALCDCLTTGAAGFAVFVFVTDGQM QVFGNVQPAVET
                                150      160      170      180      190      200

                                40      50      60      70      80      90
orfl4.pep                      GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRs
45  orfl4a                      GFFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRs
                                210      220      230      240      250      260

                                100      110      120      130      140      150
orfl4.pep                      VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG
50  orfl4a                      VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG
                                270      280      290      300      310      320

```

-135-

```

              160
orfl4.pep    RXLTNPTVSVRIMLHSG
              | | | | | | | | | |
orfl4a       RSLTNPTVSVRIMLHSGLMYSRRVVSVAKSWSFAYMPDLVSRNLRLDLPTLVX
5            330      340      350      360      370      380

```

The complete length ORF14a nucleotide sequence <SEQ ID 143> is:

	1	ATGGAGGATT	TGCAGGAAAT	CGGGTTCGAT	GTCCGCCCG	TAAAGGTAGG
	51	TCGGCAGCGC	GAACATCATC	GTCTGCATCA	TCCCAGCCC	GGCAACGGCG
10	101	AGGCGACGA	TGTATTGTTT	GCGTTCTTTT	TGGTTGGCG	CTTCGATTTT
	151	TTGCGCGTCA	TAGGGTGCGG	CGGTGTAGCC	TATCTGCCTG	ATTTTCAACA
	201	GAATGTCGGA	AAGGCGGATT	TTGCCGTCGT	CCAGACGAC	GCGGCAGCGG
	251	TGCGTGCTGT	AATTGAGGTC	GATGCGGACG	ATGCCGTCTG	TACGCAAAAG
	301	CTGCTGTTTC	ATCAGCCAGA	CGCAGGCGGC	GCAGGTGATG	CCGCCGAGTA
15	351	TTAAAACCGC	CTCGCGCGTG	CCGCCGTGGG	TTTCCACAA	GTCCGAGTGG
	401	ACTTCGGGCA	GGTCGTACAG	CGGATTGGG	TCGAGGATTT	CTTGGGGCGG
	451	CAGCTCGGTT	TTTTGCGCGT	CGGCGGTGCG	TTGTTTGTA	TAACTGCCCA
	501	AGCCCGCGTC	AATAATGCTT	TGTGCGACTG	CCTGACAACC	GGCGCAGCAG
	551	GTTTCGCGGT	CTTCGTTTTT	GTAAACGGAC	GTACAGATGCA	GGTTTTTCGG
20	601	AACGTCACAG	CCGCAGTGGG	AACAGGTTT	TTTCAATGGA	TTTCGGTTTC
	651	GTCTGTGTTT	GGTGCGGCGG	CACAATACTC	GGCAATGGCT	TTCGCGCAGT
	701	CGTCTATACC	GGTATTTTCA	GCAACGGAAA	TGCGGACGGC	GGCAATTTTT
	751	CCGCGACGGT	CGCGCCATAT	GCCCGTGTTT	TGTTCTTCAG	ACGGCAGCAG
	801	GTGCGTTTTG	TTGTACACCT	TGATGCACGG	AATATCGCCG	GCATGGATTT
25	851	CTTGCGAGTAC	GTTTTCCACG	TCTTCAATCT	GCTTGCCGCT	TCTCGGAGCG
	901	GCGGCATCGA	CGACGTGCAG	CAGCACATCG	GCTTGCGCGG	TTTCTTCCAG
	951	CGTGGCGGAA	AAGGCGGAAA	TCAGTTTGTG	CGGCAGATCG	CTGACGAATC
	1001	CGACGGTATC	GGTCAGGATA	ATGCTGCATT	CGGGACTGAT	GTACAGCCCG
	1051	CGCGCCGTCG	GTCTGAGTGT	GGCGAAAAGC	TGGTCTTTTC	CATATATGCC
	1101	CGACTTGGTC	AGCCGGTTGA	ACAGACTGGA	TTTGCCGACA	TTGGTATAG

30 This encodes a protein having amino acid sequence <SEQ ID 144>:

	1	MEDLQEIGFD	VAVKVGRQR	EHHRLHHPQP	GNGEADDVLF	AFFLVGGDFD
	51	LRVIGCGGVA	YLPDFQQNVG	KADFAVVPDD	AAAVRAVIEV	DADDAVCTQK
	101	LLFDQPDAGG	AGDAAEH*NR	LARAAGVFHK	VGLDFGQVVQ	ADLVDFGLGR
	151	QLGFLRVGGA	LFVITAAQARV	NNALCDCLTT	GAGGFAVFVF	VTDGQMQVFG
35	201	NVQPAVETGF	FHGISVSSVF	GAAQYSAMA	SRSASIPVFS	ATEMRTAAIF
	251	PAASRRMPVF	CSSDGSRSVL	LYTLMHGISP	AWISCSSTFS	SSICPLFGA
	301	AASTTCSSTS	ACAVSSSVAE	KAEISLCGRS	LTNPVTSVRI	MLHSGLMYSR
	351	RAVVSSVAKS	WSFAYMPDLV	SRLNRDLPT	LV*	

It should be noted that this sequence includes a stop codon at position 118.

40 Homology with a predicted ORF from *N.gonorrhoeae*

ORF14 shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) from *N. gonorrhoeae*:

	orf14.pep	TAGAAGXXVFVFTDSQVEVFGNIQTAVET	30
		: : : : :	
45	orf14.ng	GRQFGFFRVGGASFVITAQAGIDDALCDCLTADAAGFAVFAFVADGQMQVFGNVQPAVET	208
	orf14.pep	GGFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS	90
50	orf14.ng	GGFFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS	268
	orf14.pep	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSSVAEKAELSLCG	150
	orf14.ng	VLLYTLMHGISWAWISCSTFSTSSICPLFRAAASTTCSSTSACTVSSKVAEKAELSLCG	328
55	orf14.pep	RXLNTNPTVSVRIMLHSG	167
	orf14.ng	RSLTNTNPTVSVRIMLHAGLMYSRRVVSRAVKSWSFAYMPDLVSRNLNRLDLPTLV	382

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein having amino acid sequence <SEQ ID 146>:

-136-

1 MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
 51 LRVIGCGGVA CLPDFQQNVG EADFAVVPDD AAAVRAVIEV DADDAVCAQK
 101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
 151 QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMQVFG
 5 NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
 201 PAASRHPVVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPFLRA
 251 AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
 301 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the
 10 proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for
 vaccines or diagnostics, or for raising antibodies.

Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

15 1 ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
 51 GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTATTGTG ATGATTTTGA
 101 TGCCGAATC GGGCAGCTTC GGTTCGGCT ATGCGTCGCT GGCGGCTTGT
 151 TCGTTCGGCG CGCTGATGAT TGCGCTGTTA GACGTGTCGT CAAATATGGC
 201 GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
 251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
 20 GTGGCGGCGA TTCTGCCGTT TGTGTTTGC TATATCGGTT TGGCGAACAC
 351 CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG
 401 GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
 451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCCGC
 501 GAATCAGGAA AAAGCCAACCT GGATCGCACT CTTAAAA.CC GCGC..

25 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

1 ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL
 51 SFGALMIALI DVSSNMAMQP FKMMVGDMMV EEQKXYAYGI QSFLANTGAV
 101 VAAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK
 151 EYXPETYARY HGIDVAANQE KANWIALLKX A..

30 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
 51 AAAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
 101 CCTTTACCTT GCAAAGCTCG CAAATGAGCC GCATTTTTC AACCCTAGGC
 151 GCAGACCCGC ACAATTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG
 35 201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGGGAAGC
 251 CGCGTTTGGG CGGCCCGCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
 351 CTATGCGTCG CTGGCGGCTT TGTCGTTTCG CGCGCTGATG ATTGCGCTGT
 401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
 40 451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
 501 CTTAGCAAAT ACGGGCGCGG TCGTGCGCGG GATTCTGCCG TTTGTGTTTG
 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCCGAGACC
 601 GTGGTCGTGG CGTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
 651 GTTACGATT TTCAAAGTGA AGGAATACGA TCCGGAACCC TACGCCCGTT
 45 701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
 751 CTCTTGAAAA CCGCGCTAA GCGTTTTGGG ACGGTTACTT TGGTGCAATT
 801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
 851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCGT AGGTTATCAG
 901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
 50 951 GGTGATTGT TCGTTGTAT TGCGGAAAGT GCCGAATAAA TACCATAAGG
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
 1101 CATCGCTTGG GCGGGCATT TCACTTATCC GCTGACGATT GTGACCAACG
 1151 CTTGTTCGGG CAAGCATATG GGCACTTACT TGGGCTTGT TAACGCTCTC
 55 1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGCTCGTCC
 1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG
 1351 GTTTGA

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```

      1  MSEYTPQTAQ  QGLPALAKST  IWMLSFGFLG  VQTAFTLQSS  QMSRIFQTLG
     51  ADPHNLGWFF  ILPLAGMLV  QPIVGHYSR  TWKPRLGRR  LPYLLYGTLI
    101  AVIVMILMPN  SGSFGFGYAS  LAALSFGALM  IALLDVSSNM  AMQPFKMMVG
    151  DMVNEEQKGY  AYGIQSFLAN  TGAVVAAILP  FVFAYIGLAN  TAEKGVVPQT
    201  VVVAFYVGAA  LLVITSFTI  FKVKEYDPET  YARYHGIDVA  ANQEKANWIE
    251  LLKTAPKAFW  TVTLVQFFCW  FAFQYMWYTS  AGAIAENVWH  TTDASSVGYQ
    301  EAGNWDYVLA  AVQSVAVIC  SFVLAKVPNK  YHKAGYFGCL  ALGALGFFSV
    351  FFIGNQYALV  LSYTLIGIAW  AGIITYPLTI  VTNALSGKHM  GTYLGLFNGS
    401  ICMPIVASL  LSFVLFPLMG  GLQATMFLVG  GVVLLGAFS  VFLIKETHGG
    451  V*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N.*

15 *meningitidis*:

```

                                     10      20      30
orfl6.pep                        GHYSDRTWKPRXLXGRRLPYLLYGTLIAVIV
                                     |||||
    20  orfl6a      IFQTLGADPHSLGWFFILPLAGMLVQPIVGHYSDRTWKPRXLXGRRLPYLLYGTLIAVIV
                        50      60      70      80      90     100

                                     40      50      60      70      80      90
orfl6.pep      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI
                                     |||||
    25  orfl6a      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKGYAYGI
                        110     120     130     140     150     160

                                     100     110     120     130     140     150
orfl6.pep      QSFLANTGAVVAAILPFVFAIYIGLANTAXKGVVPQTVVVAFYVGAAALLVITSFTIFKVK
                                     |||||
    30  orfl6a      QSFLANTGAVVAAILPFVFAIYIGLANTAEKGVVPQTVVVAFYVGAAALLVITSFTIFKVK
                        170     180     190     200     210     220

                                     160     170     180
orfl6.pep      EYXPETYARYHGIDVAANQEKANWIALLKXA
                                     |||||
    35  orfl6a      EYNPETYARYHGIDVAANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI
                        230     240     250     260     270     280

    40  orfl6a      AENVVHTTDASSVGYQEAGNWDYVLAQVQSVAAVICSFVLAKVPNKYHKAGYFGCLALGA
                        290     300     310     320     330     340
  
```

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

```

      1  ATGTCGGAAT  ATACGCCTCA  AACAGCAAAA  CAAGGTTTGC  CCGCGCTGGC
     51  AAAAAGCACG  ATTTGGATGC  TCAGTTTCGG  CTTTCTCGGC  GTTCAGACGG
    101  CCTTTACCTT  GCAAAGCTCG  CAGATGAGCC  GCATCTTCCA  GACGCTCGGT
    151  GCCGATCCGC  ACAGCCTCGG  CTGTTTCTTT  ATCCTGCCGC  CGCTGGCGGG
    201  GATGCTGGTG  CAGCCGATTG  TCGGCCATTA  CTCCGACCGC  ACTTGAAGC
    251  CGCGTTTGGG  CGGCCGCCGT  CTGCCGTATC  TGCTTTATGG  CACGCTGATT
    301  GCGGTTATTG  TGATGATTTT  GATGCCGAAC  TCGGGCAGCT  TCGGTTTCGG
    351  CTATGCGTCG  CTGGCGGCTT  TGTCGTTCGG  CGCGCTGATG  ATTGCGCTGT
    401  TAGACGTGTC  GTCAAATATG  GCGATGCAGC  CGTTTAAGAT  GATGGTCGGC
    451  GACATGGTCA  ACGAGGAGCA  GAAAGGCTAC  GCCTACGGGA  TTCAAAGTTT
    501  CTTAGCGAAT  ACGGGCGCGG  TCGTGGCGGC  GATTCTGCCG  TTTGTGTTTG
    551  CGTATATCGG  TTTGGCGAAC  ACCGCCGAGA  AAGGCGTTGT  GCCGCAGACC
    601  GTGGTCGTGG  CGTTTTATGT  GGGTGCGGCG  TTGCTGGTGA  TTACCGCGC
    651  GTTCAAGATT  TTCAAAGTGA  AGGAATACAA  TCCGGAAACC  TACGCCCGTT
    701  ACCACGGCAT  CGATGTCGCC  GCGAATCAGG  AAAAAGCCAA  CTGGATCGAA
    751  CTCTTGAAAA  CCGCGCCTAA  GCGGTTTTGG  ACGGTTACTT  TGGTGCAATT
    801  CTTCTGCTGG  TTCGCCTTCC  AATATATGTG  GACTTACTCG  GCAGGCGCGA
    851  TTGCGGAAAA  CGTCTGGCAC  ACCACCGATG  CGTCTTCCGT  AGGTTATCAG
    901  GAGGCGGGTA  ACTGGTACGG  CGTTTTGGCG  GCGGTGCAGT  CGGTTGCGGC
    951  GGTGATTGT  TCGTTTGTAT  TGGCGAAAGT  GCCGAATAAA  TACCATAAGG
  
```

5
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
 1101 CATCGCTTGG GCGGGCATT TCACTTATCC GCTGACGATT GTGACCAACG
 1151 CCTGTGCGGG CAAGCATATG GGCACCTACT TGGGCCTGTT TAACGGCTCT
 1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGCTCGTCC
 1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTGA TTAAAGAAAC ACACGGCGGG
 1351 GTTTGA

This encodes a protein having amino acid sequence <SEQ ID 152>:

10
 1 MSEYTPQTAQ QGLPALAKST IWMLSFGLG VQTAFTLOSS QMSRIFQTLG
 51 ADPHSLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI
 101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
 151 DMVNEEQKGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVPQT
 201 VVVAFYVGAA LLVITSFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE
 15 251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYQ
 301 EAGNWWGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
 351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNGS
 401 ICMPOIVASL LSFVLFPMGL GLQATMFLVG GVVLLGAFS VFLIKETHGG
 451 V*

20 ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

		10	20	30	40	50	60
orfl6a.pep		MSEYTPQTAQGLPALAKSTIWMLSFGLGVQTAFTLOSSQMSRIFQTLGADPHSLGWFF					
orfl6-1		MSEYTPQTAQGLPALAKSTIWMLSFGLGVQTAFTLOSSQMSRIFQTLGADPHNLGWFF					
		10	20	30	40	50	60
orfl6a.pep		ILPPLAGMLVQPIVGHYSRTWKPRLGRRLPYLLYGTLI					
orfl6-1		ILPPLAGMLVQPIVGHYSRTWKPRLGRRLPYLLYGTLI					
		70	80	90	100	110	120
orfl6a.pep		ILPPLAGMLVQPIVGHYSRTWKPRLGRRLPYLLYGTLI					
orfl6-1		ILPPLAGMLVQPIVGHYSRTWKPRLGRRLPYLLYGTLI					
		70	80	90	100	110	120
orfl6a.pep		LAALSFGALMIALLDVSSNMAMQPFKMMVGD					
orfl6-1		LAALSFGALMIALLDVSSNMAMQPFKMMVGD					
		130	140	150	160	170	180
orfl6a.pep		LAALSFGALMIALLDVSSNMAMQPFKMMVGD					
orfl6-1		LAALSFGALMIALLDVSSNMAMQPFKMMVGD					
		130	140	150	160	170	180
orfl6a.pep		FVFAYIGLANTA					
orfl6-1		FVFAYIGLANTA					
		190	200	210	220	230	240
orfl6a.pep		FVFAYIGLANTA					
orfl6-1		FVFAYIGLANTA					
		190	200	210	220	230	240
orfl6a.pep		ANQEKANWIELLKTAPKAFWTVTLVQFFCW					
orfl6-1		ANQEKANWIELLKTAPKAFWTVTLVQFFCW					
		250	260	270	280	290	300
orfl6a.pep		ANQEKANWIELLKTAPKAFWTVTLVQFFCW					
orfl6-1		ANQEKANWIELLKTAPKAFWTVTLVQFFCW					
		250	260	270	280	290	300
orfl6a.pep		EAGNWWGVLA					
orfl6-1		EAGNWWGVLA					
		310	320	330	340	350	360
orfl6a.pep		EAGNWWGVLA					
orfl6-1		EAGNWWGVLA					
		310	320	330	340	350	360
orfl6a.pep		LSYTLIGIAWAGIITYPLTIVTNALSGKHM					
orfl6-1		LSYTLIGIAWAGIITYPLTIVTNALSGKHM					
		370	380	390	400	410	420
orfl6a.pep		LSYTLIGIAWAGIITYPLTIVTNALSGKHM					
orfl6-1		LSYTLIGIAWAGIITYPLTIVTNALSGKHM					
		370	380	390	400	410	420
orfl6a.pep		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
orfl6-1		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
		430	440	450			

Homology with a predicted ORF from *N. gonorrhoeae*

ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N.*

gonorrhoeae:

5	orf16.pep	GHYSDRTWKPR LXGRRLPYLLYGT LIAVIV	30
	orf16ng	HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYSDRTWKPR LGGRRLPYLLYGT LIAVIV	131
	orf16.pep	MILMPNSGSF GFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGD MVNEEQKXYAYGI	90
10	orf16ng	MILMPNSGSF GFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGD MVNEEQKSYAYGI	191
	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITS AFTIFKVK	150
15	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTA EKGVPQTVVVAFYVGAALLIITS AFTISKVK	251
	orf16.pep	EYXPETYARYHGIDVAANQ EKANWIAL LKXA	181
	orf16ng	EYDPETYARYHGIDVAANQ EKANWFELLKTAPKVFWTVTPVQFFCWFAFRYMW TY SAGAI	311

20 The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
51	TACTTTTCAA	ATCAAAAAAA	AGGATTACT	TTATGTCGGA	ATATACGCCT
101	CAAACAGCAA	AACAAGGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTGGAT
151	GTTGAGCTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAGCT
25	201	CGCAGATGAG	CCGCATTTT	CAAACGCTAG	GCGCAGACCC
	251	GGCTGGTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG
	301	AGTGGCTACT	ACTCAGACCG	CACTTGAAG	CCGCGCTTGG
	351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTCATC
	401	TGATGCCGAA	CTCGGCAGC	TTCGGTTTCG	GCTATGCGTC
30	451	TTGTTCGTT	CGCGCTGAT	GATTGCGCTG	TTGGACGTGT
	501	GGCGATGCAG	CCGTTTAAGA	TGATGGTCGG	CGATATGGTC
	551	AGAAAAGCTA	CGCCTACGGG	ATTCAAAGTT	TCTTAGCGAA
	601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTT	GCGTATATCG
	651	CACTGCCGAG	AAAGGCGTTG	TGCCACAAAC	CGTGGTCGTA
35	701	TGGGTGCGGC	GTTACTGATT	ATTACAGTG	CGTTCACAA
	751	AAAGAATACG	ACCCGGAAC	CTACGCCGT	TACCACGGCA
	801	CGCGAATCAG	GAAAAAGCCA	ACTGGTTCGA	ACTCTAAAA
	851	AAGTGTTTT	GACGGTTACT	CCGGTACAGT	TTTTCTGCTG
	901	CGGTATATGT	GGACTTACTC	GGCAGGCGCG	ATTGCAGAAA
40	951	CACTACCGAT	GCGTCTTCCG	TAGGCCATCA	GGAGGCGGGC
	1001	GCGTTTGGC	GCGGTGTAG		AACCGGTACG

This encodes a protein having amino acid sequence <SEQ ID 154>:

1	MIGDRRAGNH	FGFSKANTFQ	IKKKDILLYVG	IYASNSKTRF	ARAGKKHOLD
51	VELRLSRRSD	GLYPAKLADE	PHFSNARRRP	AQFGLVFHPA	AAGGDAGSAD
101	SGYSDRTWK	PR LGGRRLPY	LLYGT LIAVI	VMILMPNSGS	F GFGYASLAA
151	LSFGALMIAL	LDVSSNMAMQ	PFKMMVGD MV	NEEQSYAYG	IQSFLANTDA
201	VVAAILPFVF	AYIGLANTAE	KGVVPQTVV	AFYVGAALLI	ITSAFTISKV
251	KEYDPETYAR	YHGIDVAANQ	EKANWFELLK	TAPKVFWTVT	PVQFFCWFAF
301	RYMW TY SAGA	IAENVHHTD	ASSVGHQEAG	NRYGVLA AV*	

50 ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

		30	40	50	60	70	80
	orf16-1.pep	MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSDRT					
	orf16ng	DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGGDAGSADSGYSDRT					
55		50	60	70	80	90	100
	orf16-1.pep	WKPR LGGRRLPYLLYGT LIAVIVMILMPNSGSF GFGYASLAALSFGALMIALLDVSSNMA					
	orf16ng	WKPR LGGRRLPYLLYGT LIAVIVMILMPNSGSF GFGYASLAALSFGALMIALLDVSSNMA					
60		110	120	130	140	150	160

-140-

		150	160	170	180	190	200
5	orf16-1.pep	MQPFKMMVGDVNEEQKGYAYGIQSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTV					
	orf16ng	MQPFKMMVGDVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTV					
		170	180	190	200	210	220
10	orf16-1.pep	VVAFYVGAALLVITSFTIFKVKYEDPETYARYHGIDVAANQEKANWIELLKTAPKAFWT					
	orf16ng	VVAFYVGAALLIITSFTISKVKYEDPETYARYHGIDVAANQEKANWFELLKTAPKVFWT					
		230	240	250	260	270	280
15	orf16-1.pep	VTLVQFFCWFAFYQYMWTSAGAIENVWHTTDASSVGYQEAGNWWYGVLAQVSAVAVICS					
	orf16ng	VTPVQFFCWFAFRYMWTSAGAIENVWHTTDASSVGHQEAGNRYGVLAQVAVX					
		290	300	310	320	330	340

- 20 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

25	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGCATA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	NAAACACGTT	GNCAAAGACC	AAATCCGNGN	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
30	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AA.NTGACGG
	251	GNATTTTGAN	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TNAGGATACC
	301	CCGAGCTATG	C.TGCCACCA	AGCCCTGCCG	GTCAAACCTCG	GATCGNCTGG
	351	CAGCCAGAAT...				

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

35	1	MLFRKTTAAV	LAHTLMLNGC	TLMLWGMNNP	VSETITRKHV	XKDQIRXFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WVFNVEDSA	XXTGILXAGL	DKPFQIVXDT
	101	PSYXCHQALP	VKLGSXGSQN...			

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

40	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TGAGGATACC
45	301	CCGAGCTATG	CTCGCCACCA	AGCCCTGCCG	GTCAAACCTCG	AATCGCCTGG
	351	CAGCCAGAAT	TTCAGTACCG	AAGGCCTTTG	CCTGCGCTAC	GATACCGACA
	401	AGCCTGCCGA	CATCGCCAAG	CTGAAACAGC	TCGGGTTTGA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAAATA
	501	CTACGCCACA	CCGCAAAAAC	TGAACGCCGA	TTACCATTIT	GAGCAAAGTG
50	551	TGCCTGCCGA	TATTTATTAC	ACGGTTACTG	AAGAACATAC	CGACAAATCC
	601	AAGCTGTTTG	CAAATATCTT	ATATACGCCC	CCCTTTTTGA	TACTGGATGC
	651	GGCGGGCGCG	GTACTGGCCT	TGCCTGCGGC	GGCTCTGGGT	GCGGTCGTGG
	701	ATGCCGCCCG	CAAATGA			

This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

55	1	MLFRKTTAAV	LAATLMLNGC	TLMLWGMNNP	VSETITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WVFNVEDSA	KLTGILKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKLESPGSQN	FSTEGLCRLY	DTDKPADIAC	LKQLGFCAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIIY	TVTEEHTDKS

Homology with a predicted ORF from *N.meningitidis* (strain A)

5 *meningitidis:*

20 The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

This encodes a protein having amino acid sequence <SEQ ID 160>:

ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:

BNSDOCID: <WO 9924578A2_1_>

ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.*

	orf28.pep	MLFRKTTAAVLAHTLMLNGCTMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK	60
	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSTITRKHVDDKQIRAFGVVAEDNAQLEK	60
15	orf28.pep	GSLVMMGGKGYWFVVPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXGSQN	120
	orf28ng	GSLVMMGGKGYWFAVNPEDSAKITGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN	120

20	1	ATGTTGTTCC	GTA AACGAC	CGCCGCCGTT	TTGGCGGC AA	CCTTGATACT
	51	GAACGGCTGT	ACGATGATGT	TGCGGGGGAT	GAACAACCCG	GTCAGCCAAA
	101	CAATACCCCG	CAAACACGTT	GACAAGACCC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGACGCTTGG	TGATGATGGG
	201	CGGGAATAAC	TGGTTCGCGC	TCAATCCCGA	AGATTCCGCG	AAGCTGACGG
25	251	GCCTTTTGAA	GGCCGGGTTG	GACAAGCCCT	TCCAATAATG	TGAGGATAACC
	301	CCGAGCTATG	CCCGCCACCA	AGCCCTGCCG	GTCAAAATCG	AAGCGCCCGG
	351	CAGCCAGAAT	TTCAGTACCG	GAGGTCTTTG	CCTGCCTAT	GATACCGGCA
	401	GACCTGACGA	CATCGCCAAG	CTGAACAGC	TTGAGTTTAA	AGCGGTCAAA
	451	CTCGACAAT	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAAATA
30	501	CTACGCACG	CCGCAAAAAC	TGAACGCCGA	TTATCATTTT	GAGCAAAGTG
	551	TGCCCGCCGA	TATTTATTAT	ACGGTTACTG	AAAAACATAC	CGACAAATCC
	601	AAGCTGTTTG	GAAATATCTT	ATATACGCCC	CCCTTGTTGA	TATTGGATGC
	651	GGCGGCCGCG	GTGCTGGTCT	TGCCTATGGC	TCTGATTGCA	GCCGCGAATT
	701	CCTCAGACAA	ATGA			

35

1	MLFRKTTAAV	LAATLLILNGC	TMMLRGMNNP	VSQTITRKHV	DKDQIRAFGV
51	VAEDNAQLEK	GSLVMGGKY	WFAVNPEDSA	KLTGLLKAGL	DKPFQIVEDT
101	PSYARHQALP	VKFEAPGSQN	FSTGGLCLRY	DTGRPADIAY	LKQLEFKAVK
151	LDNRITTYTRC	VSAGKYYAT	PQKLNADYHF	EQSPVDIAK	VTTEKHTDKS
201	KLFGNILYTP	PLLILDAAAA	VLVLPMALIA	AANSSDK*	

		10	20	30	40	50	60
	orf28-1.pep	MLFRKTTAAVLAATLMLNGCTLMWGMNPNVSETITRKHVDKQIRAFGVVAEDNAQLEK					
		: : : : : : :					
45	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf28-1.pep	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSON					
		: : : : : : :					
50	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf28-1.pep	FSTEGLCRLRYDTDKPADIAKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		: : : : :					
55	orf28ng	FSTGGLCRLRYDTGRPDDIAKLKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		130	140	150	160	170	180
		190	200	210	220	230	239
60	orf28-1.pep	EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
		: : : : : : :					
	orf28ng	EQSVPADIYYTVTEKHTDKSKLFGNILYTPPLLILDAAAAVLVLPMALIAAANSSDKX					

190 200 210 220 230

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

```

1   ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
15  51  TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
    101 CGTTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTTCA CGGCGGTGTA
    151 GACGGCGGTT TTAAGTTTCA CCAACTTCAT CGAACATGGT CGGAAATCCA
    201 TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCCG
    251 GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
20  301 ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
    351 AGAAAATGCC GGTGCCGCCT CTGGT..

```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```

1   ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHDS KSTSDFSGGV
25  51  DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYYVKGTSTK
    101 TKTSIVPQAP FSDRWLEENA GAASG..

```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

```

1   ATGAATTTGC CTATTCAAAA ATTCATGATG CTGTTTGAG CAGCAATATC
30  51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
    101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
    151 TTTGGTAATG CTCGCGGCAG TGTTAAAAAG CGGGTTTACG CCGTCCAGAC
    201 ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
    251 CAGGGTTTGA AGGTGTTATC GGTTATGAAA CCCATTTTTC AGGGCACGGA
    301 CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
    351 TTTCAGCGGC GGTGTAGACG GCGGTTTTC TGTTTACCAA CTTTCATCGAA
35  401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
    451 GATTATCCGC CCCCCGGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
    501 AGGAACTTCA ACAAACAACA AGACTAATAT TGTCCTCAA GCCCATTTT
    551 CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
    601 CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
40  651 TTGGTGGGCT AACCCTATGG ATGATGTTTC CGGCATCGTC CAAGGTGCGG
    701 TTAATCCTTT TTTAATGGGT TTCAAGGAG TAGGGATTGG GGCAATTACA
    751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
    801 AGGTATTAAT GATTTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG
    851 CGAGCCTATT ACAGGACAGT GCTTTTGGCG TAAAAGACGG TATCAACTCT
45  901 GCCAAACAAT GGGCTGATGC CCATCCAAAT ATAACAGCTA CTGCCCAAAC
    951 TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTTGGAGA GGTAAAAAAG
1001 TAGAACTTAA CCCGACTAAA TGGGATTGGG TTAATAATAC CGGTTATAAA
1051 AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
1101 GAATAAACCT ATTAATCTT TACCAAACAG TGCCGCTGAA AAAAGAAAAC
50 1151 AAAATTTTGA GAAGTTTAAT AGTAACTGGA GTTCAGCAAG TTTTGATTCA

```

5
1201 GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA
1251 AGTTAAAACT CGATACACTA GTTTAGATGG AAAAATTACA ATTATAAAAG
1301 ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCACG AAAACAGTAT
1351 CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAAATTTAC AAGGTAAGCA
1401 AGCAAAAGAT TATTTACAAC AACAAACTCA TATCAGGAAC TTAGACAAAT
1451 GA

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

10
15
1
MNLPIQKFM LFAAAISLLO IPISHANGLD ARLRDDMOAK HYEPGGKYHL
51 FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG
101 HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151 DYPPPGGARD IYSYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS
201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT
251 DSAVSPVTD TAAQOTLQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS
301 AKQWADAHNP ITATAQTALS AEAAGTVWR GKKVELNPTK WDWVKNTGYK
15 351 KPAARHMOTL DGEMAGGNKP IKSLPNSAAE KRKQNFELFN SNWSSASFDS
401 VHKTLPNAP GILSPDKVKT RYTSLDGKIT IIKDNENNYF RIHDNSRKQY
451 LDSNGNAVKT GNLOGKQAKD YLQQQTHIRN LDK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

20 ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis*:

				10	20	30
	orf29.pep			VSPVLPITHERTGFEGVIGYETHFSGHGHE		
25	orf29a	EPGGKYHLFGNARGSVKNRVYAVQTFDATAVGPILPITHERTGFEGIIGYETHFSGHGHE				
		50 60 70 80 90 100				
			40 50 60 70 80 90			
30	orf29.pep	VHSPFDHHDHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDGYDGPQAAAXYPPPGGARDIY				
	orf29a	VHSPFDHHDHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIY				
		110 120 130 140 150 160				
			100 110 120			
35	orf29.pep	SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG				
	orf29a	XXYVVKGTSTKTKSNIVPRAPFSDRWLKENAGAASGFFSRADAGKLIWESDPNKNWWANR				
		170 180 190 200 210 220				
40	orf29a	MDDIRGIVQGAVNPFLMGFQGVGIGAITDSAVSPVTD TAAQOTLQGXNHLGXLSPEAQLA				
		230 240 250 260 270 280				

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

45
50
55
60
1
ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGACG CAGCAATATC
51 GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
151 TTTGGTAATG CTCGCGGCAG TGTAAAAAT CGGGTTTACG CCGTCCAAAC
201 ATTTGATGCA ACTGCGGTCG GCCCATACT GCCTATTACA CACGAACGGA
251 CAGGATTTGA AGGCATTATC GGTATGAAA CCCATTTTTC AGGACATGGA
301 CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACCTTCTGA
50 351 TTTGAGCGGC GCGGTAGACG GTGGTTTAC CGTTTACCAA CTTTCATCGGA
401 CAGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAGGCAGC
451 GATTATCCGC CCCCCGAGG AGCAAGGGAT ATATACANNT ANTATGTCAA
501 AGGAACTTCA ACAAAAACAA AGAGTAATAT TGTTCCCCGA GCCCATTTT
55 551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCTCTGG TTTTTCAGC
601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
651 TTGGTGGGCT AACGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG
701 TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
801 AGGTATNAAT CATTTAGGAA ANTAAAGTCC CGAAGCACAA CTTGCGGCTG
60 851 CAACCGCATT ACAAGACAGT GCTTTTGGCG TAAAAGACGG TATCAATTCC
901 GCCAGACAA GGGCTGATGC CCATCCGAAT ATAAGTGCAA CAGCCCAAC

-145-

5
10

```

951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTTTGGGGC GGTAACAAAAG
1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TTAAAAATAC NGGCTATAAN
1051 ACACCTGCTG TTCGCACCAT GCATACTTTG GATGGGGAAA TGGCCGGTGG
1101 GAATAGACCG CCTAAATCTA TAACGTCCAA CAGCAAAGCA GATGCTTCCA
1151 CACAACCGTC TTTACAAGCG CAACTAATTG GAGAACAAAT TANNNNNGGG
1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTTA CGGATTTAAA
1251 TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAAT ATTGTTAGCC
1301 ATCCANCAAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT
1351 NATAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG
1401 TACAGCATTT AGACCAACAT CAGGTAAAAA ATATTATGAT GATTTATAG

```

This encodes a protein having amino acid sequence <SEQ ID 168>:

15
20

```

1 MNXPIQKFMM LFAAAISXLO IPISHANGLD ARLRDDMQAK HYPEGGKYHL
51 FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG
101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151 DYPPPGGARD IYXXYVKGTS TTKSNIVPR APFSDRWLKE NAGAASGFFS
201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT
251 DSAVSPVTD TAAQTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGINS
301 ARQWADAHPN ITATAQTALA VAXAATTVWG GKKVELNPTK WDWVKNTRYX
351 TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGEQIXXG
401 HAYNKHVIRQ QEFTDLNINS PADFARHIEN IVSHPXNMKE LPRGRTAYWD
451 XKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL*

```

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

25
30
35
40
45
50
55
60

```

          10      20      30      40      50      60
orf29a.pep MNXPIQKFMM LFAAAISXLO IPISHANGLD ARLRDDMQAK HYPEGGKYHL FGNARGSVKN
          10      20      30      40      50      60
orf29-1     MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYPEGGKYHL FGNARGSVKK

          70      80      90     100     110     120
orf29a.pep RVYAVQTFD TAVGPILPIT HERTGFEGII GYETHFSGHG HEVHSPFDNH DSKSTSDFSG
          70      80      90     100     110     120
orf29-1     RVYAVQTFD TAVSPVLPIT HERTGFEGV GYETHFSGHG HEVHSPFDH DSKSTSDFSG

          130     140     150     160     170     180
orf29a.pep GVDGGFTVYQL HRTGSEIHP EDGYDGPQGS DYPPPGGARDI YXXYVKGTSTK TKSNI VPR
          130     140     150     160     170     180
orf29-1     GVDGGFTVYQL HRTGSEIHP EDGYDGPQGS DYPPPGGARDI YSYVKGTSKT KTNIV PQ

          190     200     210     220     230     240
orf29a.pep APFSDRWLKEN AGAASGFFS RADEAGKLIW ESDPNKNWWAN RMDDIRGIV QGAVNPFLMG
          190     200     210     220     230     240
orf29-1     APFSDRWLKEN AGAASGFFS RADEAGKLIW ESDPNKNWWAN RMDVIRGIV QGAVNPFLMG

          250     260     270     280     290     300
orf29a.pep FQGVGIGAITD SAVSPVTD TAAQTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGINS
          250     260     270     280     290     300
orf29-1     FQGVGIGAITD SAVSPVTD TAAQTLQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS

          310     320     330     340     350     360
orf29a.pep ARQWADAHPN ITATAQTAL AVAXAATTVWGG KVELNPTKWD WVKNTRYXTP AVRTMHTL
          310     320     330     340     350     360
orf29-1     AKQWADAHPN ITATAQTAL SAAEAGTVWRG KVELNPTKWD WVKNTRYXTP ARHMQTL

          370     380     390     400     410     420
orf29a.pep DGEMAGGNRP PKSITSNSK DASTQPSLQA QLIGEQIXXG HAYNKHVIRQ QEFTDLNINS
          370     380     390     400     410
orf29-1     DGEMAGGNRP KISLP-NSAAEKRQN FEKFNSNWSSAS FDSVHKTLP NAPGILSPDKVK

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N. gonorrhoeae*:

5	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSGHGHE	30
	orf29ng	EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE	102
10	orf29.pep	VHSPFDHHDSDKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDYDGPQAAXYPPPGGARDIY	90
	orf29ng	VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGGYPPPGGARDIY	162
	orf29.pep	SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG	125
	orf29ng	SYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANR	222

15 The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

	1	MNLPIQKFMM	LFAAAISLLQ	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
	51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG
20	101	HEVHSPFDNH	DSKSTSDFSG	GVDGGFTVYQ	LHRTGSEIHP	EDGYDGPQGG
	151	GYPPPGGARD	IYSYHIKGT	TKTKINTVPQ	APFSDRWLKE	NAGAASGFLS
	201	RADEAGKLIW	ENDPDKNWR	NRMDDIRGIV	QGAVNPFITG	FQGLGVGAI
	251	DSAVSPVTYA	AARKTLQGIH	NLGNLSPEAQ	LAAATALQDS	AFAVKDSINS
	301	ARQWADAHPN	ITATAQTALA	VTEAATTWVG	GKKVELNPAK	WDWVKNTGYK
	351	KPAARHMQTV	DGEMAGGNKP	LESKNVTVTN	NFFENTGYTE	KVLRQASNGD
25	401	YHGFPSQVDA	FSENGTVIQI	VGGDNIVRHK	LYIPGSYKKG	DGNFEYIREA
	451	DGKINHRLFV	PNQQLPEK*			

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

	1	atgAATTTGC	CTATTCAAAA	ATTCATGATG	ctgttggcAg	cggcaatatac
30	51	gatgctGCat	ATCCCCATTA	GTCATGCGAA	CGGTTTGGAT	GCCCCGTTTGC
	101	GCGATGATAT	GCAGGCAAAA	CACTACGAAC	CGGGTGGCAA	ATACCATCTG
	151	TTTGGTAATG	CTCGCGGCAG	TGTTAAAAAT	CGGGTTTGCG	CCGTCCAAAC
	201	ATTTGATGCA	ACTGCGGTCG	CCCCATACT	GCCTATTACA	CACGAACGGA
	251	CAGGATTTGA	AGGTGTTATC	GGCTATGAAA	CCCATTTTTC	AGGACACGGA
35	301	CACGAAGTAC	ACAGTCCGTT	CGATAATCAT	GATTCAAAAA	GCACCTCTGA
	351	TTTCAGCGGC	GGCGTAGACG	GCGGTTTTAC	CGTTTACCAA	CTTCATCGGA
	401	CAGGTCGGA	AATACATCCC	GCAGACGGAT	ATGACGGGCC	TCAAGGCGGC
	451	GGTTATCCGG	AACCACAAGG	GGCAAGGGAT	ATATACAGCT	ACCATATCAA
	501	AGGAACTTCA	ACCAAACAA	AGATAAACAC	TGTTCCGCAA	GCCCCTTTTT
	551	CAGACCGCTG	GCTAAAAGAA	AATGCCGGTG	CCGCTTCCGG	TTTTCTCAGC
40	601	CGTGCGGATG	AAGCAGGAAA	ACTGATATGG	GAAAACGACC	CCGATAAAAA
	651	TTGGCGGGCT	AACCGTATGG	ATGATATTCG	CGGCATCGTC	CAAGGTGCGG
	701	TTAATCCTTT	TTTAAACGGG	TTTCAAGGGG	TAGGGATTGG	GGCAATTACA
	751	GACAGTGCGG	TAAGCCCGGT	CACAGATACA	GCCGCTCAGC	AGACTCTACA
45	801	AGGTATTAAT	GATTTAGGAA	ATTAAAGTCC	GGAAGCACAA	CTTGCCCGCG
	851	CGAGCCTATT	ACAGGACAGT	GCCTTTGCGG	TAAAAGACGG	CATCAATTCC
	901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAACAGCAA	CAGCCCAAAC
	951	TGCCCTTGCC	GTCAGCAGAG	CCGCAGGTAC	GGTTTGGCGC	GGTAAAAAAG
	1001	TAGAACTTAA	CCCGACCAAA	TGGGATTGGG	TTAAAAATAC	CGGCTATAAA
50	1051	AAACCTGCTG	CCCGCCATAT	GCAGACTGTA	GATGGGGAGA	TGGCAGGGGG
	1101	GAATAGACCG	CCTAAATCTA	TAACGTCGGA	AGGAAAAGCT	AATGCTGCAA
	1151	CCTATCCTAA	GTTGGTTAAT	CAGCTAAATG	AGCAAAACTT	AAATAACATT
	1201	GCGGCTCAAG	ATCCAAGATT	GAGTCTAGCT	ATTCATGAGG	GTAAAAAATA
	1251	TTTTCCAATA	GGAAGTCAA	CTTATGAAGA	GGCAGATAGA	CTAGGTAAAA
	1301	TTTGGGTGG	TGAGGTGCA	AGACAACTA	GTGGAGGCGG	ATGGTTAAGT
55	1351	AGAGATGGCA	CTCGACAATA	TCGGCCACCA	ACAGAAAAAA	AATCACAATT
	1401	TGCAACTACA	GGTATTCAAG	CAAATTTTGA	AACTTATACT	ATTGATTCAA
	1451	ATGAAAAAAG	AAATAAAATT	AAAAATGCAC	ATTTAAATAT	TAGGTAA

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

60	1	MNLPIQKFMM	LLAAAISMLH	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
	51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG

101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG
 151 GYPEPQGARD IYSYHIKGTSTKTKINTVPQ APFSDRWLKE NAGAASGFSL
 201 RADEAGKLIW ENDPDKNWRANRMDDIRGIV QGAVNPFLTG FQGVGIGAIT
 251 DSAVSPVTD TAAQQTLOGIN DLGNLSPEAQ LAAASLLQDS AFAVKDGIN
 301 ARQWADAHPN ITATAQTALA VAEAAGTVWR GKKVELNPTK WDWVKNTGYK
 351 KPAARHMQTV DGEMAGGNRP PKSITSEGKA NAATYPKLVN QLNEQNLNNI
 401 AAQDPRLSLA IHEGKKNFPI GTATYEEADR LGKIWVGEGA RQTSGGGWLS
 451 RDGTRQYRPP TEKKSQFATT GIQANFETYT IDSNEKRNKI KNGHLNIR*

ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

10	orf29ng-1.pep	10	20	30	40	50	60
	MNLPIQKFMMLLAAAI	SMLHIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN					
	orf29-1	MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK					
15	orf29ng-1.pep	70	80	90	100	110	120
	RVCAVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG						
	orf29-1	RVYAVQTFDATAVSPVLPITHERTGFEGVIGYETHFSGHGHEVHSPFDHHDHDSKSTSDFSG					
20	orf29ng-1.pep	130	140	150	160	170	180
	GVDGGFTVYQLHRTGSEIHPADGYDGPQGGGYPEPQGARDIYSYHIKGTSTKTKINTVPQ						
	orf29-1	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGSDDYPPGGARDIYSYVKGSTSTKTKTNIVPQ					
25	orf29ng-1.pep	190	200	210	220	230	240
	APFSDRWLKENAGAASGFSLRADEAGKLIWENDPDKNWRANRMDDIRGIVQGAVNPFLTG						
	orf29-1	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG					
30	orf29ng-1.pep	250	260	270	280	290	300
	FQGVGIGAITDSAVSPVTD TAAQQTLOGINDLGNLSPEAQ LAAASLLQDS AFAVKDGIN						
	orf29-1	FQGVGIGAITDSAVSPVTD TAAQQTLOGINDLGKLSPEAQ LAAASLLQDS AFAVKDGIN					
35	orf29ng-1.pep	310	320	330	340	350	360
	ARQWADAHPNITATAQTALAVAEAAGTVWRGKKVELNPTKWDWVKNTGYKKPAARHMQTV						
	orf29-1	AKQWADAHPNITATAQTALSAAEAAGTVWRGKKVELNPTKWDWVKNTGYKKPAARHMQTL					
40	orf29ng-1.pep	370	380	390	400	410	419
	DGEMAGGNRPKPSI-TSEGKANAATYPKLVNQLNEQNLNNIAAQDPRLSLAIHEGKKNFP						
	orf29-1	DGEMAGGNKPIKSLPNSAAEKRKQNFEFKNSNWSSASFSDSVHKTLPNAPGILSPDKVKT					
45	orf29ng-1.pep	420	430	440	450	460	470
	IGTATYEEADRLGKIWVGEGARQTSGGGWLSRDGTRQYRPPTEKKSQFATTGIQANFETY						
50	orf29-1	RYTSLDGKITIIKDNENNYFRIHDNSRKQYLDNNGNAVKTGNLQKGQAKDYLQQQTHIRN					
55	orf29ng-1.pep	430	440	450	460	470	480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 21

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

-148-

```

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
101 ACACGCGGGC AGATGCACCG ATGCAG...

```

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

```

5      1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

```

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAGGAG
151 ATGAAGGAGA CAGAGGGGGC GTTCTCTCCA TTGGCTATCT TGGGTGGTGC
201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
15 401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

```

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

```

20      1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
51 MKETEGAFLP LAILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
101 PGGVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
151 DNTGKTLPGQ GIGRHPWES KSTDRSWKNR F*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

```

      10      20      30      40
orf30.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
30 orf30a   MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
      10      20      30      40      50      60
orf30a     LXILGGAAIGMWTQHGFSAATTGRPASVRDVIAAGGLGAIPGXVGAAGKVVVSFAKYGREI
      70      80      90     100     110     120

```

35 The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

```

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAGGAG
40 151 ATGAAGGANA CAGNGGGGGC GTTCTCTCCA TTGGNTATCT TGGGTGGTGC
201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
45 401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 178>:

```

50      1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
51 MKXTXGAFLP LXILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
101 PGXVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
151 DNTGKTLPGQ GIGRHPWES KSTDRSWKNR F*

```

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

```

orf30a.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60

```

-149-

```

      |||
orf30-1  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP  60
      |||
5  orf30a.pep  LXILGGAAGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI  120
      |||
orf30-1  LAILGGAAGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI  120
      |||
10 orf30a.pep  KIGNNMRIAPFGNRTGHPIGKFFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR  180
      |||
orf30-1  KIGNNMRIAPFGNRTGHPIGKFFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR  180
      |||
orf30a.pep  FX
      ||
15 orf30-1  FX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae*:

```

20 orf30.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ  42
      |||
orf30ng  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP  60

```

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

```

25      1  ATGAAAAAAC  AAATCACCGC  AGCCGTAATG  ATGCTGTCTA  TGATCGCCCC
      51  CGCAATGGCA  AACGGATTGG  ACAATCAGGC  ATTGAAGAC  CAAGTGTTC
101  101  ACACGCGGGC  AGATGCGCCG  ATGCAGTTGG  CCGAGCTTTC  TCAGAAGGAG
      151  ATGAAGGAGA  CTGAAGGGGC  TTTTCTTCCA  TTGGCTATCT  TGGGTGGTGC
201  201  TGCCATTGGT  ATGTGGACAC  AGCATGGTTT  TAGTTATGCA  ACGACAGGCA
      251  GACCAGCTTC  TGTTAGAGAT  GTTGCTGGCG  GATTAGGCGC  AATTCCTGGT
301  301  GATGTAGGTG  CTGCAGGAAA  GGTGTGTTCC  TTTGCTAAAT  ATGGACGTGA
30      351  GATTAAATC  GGCAATAATA  TGCGGATAGC  CCCTTCGGT  AATAGAACAG
      401  GTCATCCTAT  TGGAAAATTT  CCCCATTATC  ATCGTCGAGT  TACGGATAAT
      451  ACGGGCAAGA  CTTTGCCTGG  ACAGGAATT  GGTCTCATC  GCCCTTGGGA
      501  ATCAAAATCT  ACGGACAGAT  CATGGAAAA  CCGCTTCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 180>:

```

35      1  MKKQITAAVM  MLSMIAPAMA  NGLDNQAFED  QVFHTRADAP  MQLAELSQKE
      51  MKETEGAFLP  LAILGGAAG  MWTQHGFSA  TTGRPASVRD  VAGGLGAIPG
101  101  DVGAAGKVVS  FAKYGREIKI  GNNMRIAPFG  NRTGHPIGKF  PHYHRRVTDN
      151  TGKTLPGQGI  GRHRPWESKS  TDRSWKNRF*

```

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

```

40      10      20      30      40      50      60
orf30ng.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
      |||
orf30-1  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
      10      20      30      40      50      60
45      70      80      90      100     110
orf30ng.pep  LAILGGAAGMWTQHGFYSYATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI
      |||
orf30-1  LAILGGAAGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI
      70      80      90      100     110     120
50      120     130     140     150     160     170
orf30ng.pep  KIGNNMRIAPFGNRTGHPIGKFFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
      |||
55 orf30-1  KIGNNMRIAPFGNRTGHPIGKFFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
      130     140     150     160     170     180
60      180
orf30ng.pep  FX
      ||
orf30-1  FX

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 22

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

```

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
51 GrTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151 GCACCTGTTT GTg.CGTTaC AAATATCTTT TCTTTTCTT TATTGGGCTT
10 201 TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA
251 TT..

```

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

```

1 MNKTLYRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVFPFGTTH
51 APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..

```

15 Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

```

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
51 GGTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
20 151 GCACCTGTTT GTCGTTCAAA TATCTTTTCT TTTTCTTTAT TGGGCTTTTC
201 TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGCT GATGGCATT..

```

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

```

1 MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPFGTTH
51 APVCRSNIFF SLLGFSLCL AVGTANIAFA DGI..

```

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N.gonorrhoeae*:

```

30 orf31.pep MNKTLYRVIFNRKRGAVXAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF 60
   |||||
orf31ng MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAF 54

orf31.pep SFSLLGFSLCLAVGTXNIAFADGI 84
   ||
orf31ng CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNIPQVNIQTPTSAGVSV 114

```

35 The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

```

1 ATGAACAAAA CCCTCTATCG TGTGATTTTC AACCGCAAAC GCGGTGCTGT
51 GGTAGCTGTT GCCGAAACCA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGGTTCGGG CAGCGTTTAT GTGAAATCCG TTTCTTTCAT TCCTACTCAT
40 151 TCCAAAGCCT TTTGTTTTTC TGCATTAGGC TTTTCTTTAT GTTTGGCTTT
201 GGGTACGGTC AATATTGCTT TTGCTGACGG CATTATTACT GATAAAGCTG
251 CTCCTAAAAC CCAACAAGCC ACGATTCTGC AAACAGGTaa cGGCATAACCG
301 CAAGTCAATA TTCAAACCCC TACTTCGGCA GGGGTTTCTG TTAATCAATA
351 TGCCCAAGTT GATGTGGGTA ATCGCGGGGC GATTTTAAAC AACAGTCGCA
401 GCAACACCCA AACACAGCTA GGCGGTTGGA TTCAAGGCAA TCCTTGTTTG
45 451 ACAAGGGGCG AAGCACGTGT GGTGTGTAAC CAAATCAACA GCAGCCATCC
501 TTCACAACTG AATGGCTATA TTGAAGTGGG TGGACGACGT GCAGAAGTCG
551 TTATTGCCAA TCCGGCAGGG ATTGCAGTCA ATGGTGGTGG TTTTATCAAT
601 GCTTCCCGTG CCACTTTGAC GACAGGCCAA CCGCAATATC AAGCAGGAGA
651 CTTTAGCGGC TTTAAGATAA GGCAAGGCAA TGCTGTAATC GCCGGACACG

```

-151-

701 GTTTGGATGC CCGTGATACC GATTTCACAC GTATTCTTGT ATGCCAACAA
 751 AATCACCTTG ATCAGTACGG CCGAACAAAGC AGGCATTCGT AA

This encodes a protein having amino acid sequence <SEQ ID 186>:

5 1 MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
 51 SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP
 101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
 151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
 201 ASRATLTGQ PQYQAGDFSG FKIRQNAVI AGHGLDARDT DFTRILVCQ
 251 NHLDQYGRS RHS*

10 This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein from *Erwinia chrysanthemi* (accession number L39897):

 orf31ng 96 GNGIPQVNIQTPTSAGVSVNQAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE 154
 GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L
 HecA 45 GNGVPVNIATPDASGLSHNRYHDFVNDNRGLILNNGTARLTSPQLGGLIQNPNLNGRA 104
 15 Orf31ng 155 ARVVVNQINSSHPSQLNGYIEVGRRRAEVVIANPAGIAVNGGGFINASRATLTGQPOQYQ 214
 A ++N++ S + S+L GY+EV G+ A VV+ANP GI +G GF+N R TLTTG PQ+
 HecA 105 AAAILNEVSPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTPQFD 164
 20 Orf31ng 215 -AGDFSGFKIRQNAVIAGHGLDARDTDF 242
 AG SG +R G+ +I G GLDA +D+
 HecA 165 AAGGLSGLDVRGGDILIDGAGLDASRSDY 193

Furthermore, ORF31ng and ORF31-1 show 79.5% identity in 83 aa overlap:

25 10 20 30 40 50 60
 orf31-1.pep MNKTLRVIFNRKRGAVVAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNI
 |||||:||||| |||:||||| |||:|
 orf31ng MNKTLRVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYKSVSFIPTH-----SKAFC
 10 20 30 40 50
 30 70 80
 orf31-1.pep FSLGFSCLAVGTANIAFADGI
 || |||||:|||||
 orf31ng FSALGFSCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSVN
 60 70 80 90 100 110

35 On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 23

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 187>:

40 1 ATGAATACTC CTCCTTTTGT CTGTGGATT TTTTGCAAGG TCATCGACAA
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
 151 GCGCTTGCC CTGATTTGCC CGATGTTCCC TCGGTCATC AGGATATCA
 201 TGCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCG..

45 This corresponds to the amino acid sequence <SEQ ID 188; ORF32>:

 1 MNTPPFVCWI FCKVIDNEGD IGVSRLARV LHRELGWQVH LWTDDVSALR
 51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT A..

Further work revealed the complete nucleotide sequence <SEQ ID 189>:

50 1 ATGAATACTC CTCCTTTTGT CTGTGGATT TTTTGCAAGG TCATCGACAA
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT

-152-

5
10
15
20

```

151 GCGCTTTGCC CTGATTGGCC CGATGTTCCC TGC GTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
351 CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
401 GTGTTCAAAA ATATTTTGGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
501 CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
651 CAGCCTCAA CAAAGCGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TCGCGGCCCA GCTTGGGGGC AAACCTTCTT
851 TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAATCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTCCGC
951 ACACGCGCGT TTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAACCTTGC AACAAATCA AAACGGCTGG
1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTACG
1101 TCCTGAAAAA CTCGCTGCCT TTGTTTCAA GCATCAAAA ATACGCTAG

```

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

25

```

1 MNTPPFVCWI FCKVIDNFGD IGVSRLARV LHRELGWQVH LWTDDVSALR
51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
151 LIRERDYCEA VRFDTEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
201 QAGSPMTLLL AGTQIIDSLK QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATQRLCEW QTLQQHQNGW
351 RQGAEDWSRY LFGQPSAPEK LAAEVSKHQK IR*w

```

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N. meningitidis*:

35

```

      10      20      30      40      50      60
orf32.pep MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
          ||||| ||||| ||||| ||||| ||||| |||||
orf32a    MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
          10      20      30      40      50      60

40      70      80
orf32.pep CVHQDIHVRTWHSDAADIDTA
          ||||| ||||| ||||| ||||| ||||| |||||
orf32a    CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
          70      80      90      100     110     120

```

45 The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

50
55
60

```

1 ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
51 TTTCCGGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGTT TTGCACCGCG
101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151 GCGCTTTGCC CTGATTGGCC CGATGTTTNC TGC GTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 CTGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
351 CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
401 GTGTTCNAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
451 CTGATACGCG AACCGGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
501 CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNNGGGGCG ANATTATCGA
651 CAGCCTCAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT

```

-153-

5
 801 CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCTTCT
 851 TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC
 901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
 951 ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
 1001 CACAACGCCT CGAATGTTGG CAAATCCTGC AACAACATCA AAACGGCTGG
 1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC
 1101 ATCCGAAAAA CTCGCCGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG

This encodes a protein having amino acid sequence <SEQ ID 192>:

10
 1 MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
 51 ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL
 101 HIIRRHKPLW LXWEYLSAEX SNERLHXMP SNERLHXMP PQESVXKXFW FMGFSEXSGG
 151 LIRERDYCEA VRFDGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMR
 201 QAGSPLTLL AGAXIIDSLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV
 251 PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
 15 301 AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLCW QILQQHONGW
 351 RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

20
 orf32-1.pep 10 20 30 40 50 60
 MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
 orf32a 10 20 30 40 50 60
 MNTPPFSAGXFCCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
 25
 orf32-1.pep 70 80 90 100 110 120
 CVHQDIHVRTWHSDAADIDTAPVDPVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAEE
 orf32a 70 80 90 100 110 120
 CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
 30
 orf32-1.pep 130 140 150 160 170 180
 SNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS
 orf32a 130 140 150 160 170 180
 SNERLHXMPSPQESVXKXFWFMGFSEXSGGLIRERDYCEAVRFDGALRKRLMLPEKNXP
 35
 orf32-1.pep 190 200 210 220 230 240
 EWLLFGYRSDVWAKWLEMRQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDVFQTA
 orf32a 190 200 210 220 230 240
 EWLLFGYRSDVWAKWLEMRQAGSPLTLLLAGAXIIDSLKQNGVIPQDALQNDGDVFQTA
 40
 orf32-1.pep 250 260 270 280 290 300
 SVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH
 orf32a 250 260 270 280 290 300
 SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH
 45
 orf32-1.pep 310 320 330 340 350 360
 AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLCWQTLQQHONGWRQGAEDWSRY
 orf32a 310 320 330 340 350 360
 AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATQRLCWQILQQHONGWRQGAEDWSRY
 50
 orf32-1.pep 370 380
 LFGQPSASEKLAAFVSKHQKIRX
 orf32a 370 380
 LFGQPSASEKLAAFVSKHQKIRX

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N.gonorrhoeae*:

orf32.pep MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP 57
 ||| | |||||

-154-

orf32ng MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLP 60
 orf32.pep DVPCVHQDIHVRTWHSDAADIDTA 81
 ||| |||||
 5 orf32ng DVPFVHQDIHVRTWHSDAADIDTAPVDPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

10
 1 MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDVVS
 51 ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVDAV IETFACDLPE
 101 NVLNIIRRHK PLWLNWEYLS AEESENERLHL MPSPQEGVQK YFWFMGFSEK
 151 SGGIIRERDY REAVRFDTEA LRRRLVLP EK NAPEWLLFGY RGDVWAKWLD
 201 MWQQAGSLMT LLLAGAQIID SLKQSGVPIQ NALQNEGGVF QTASVRLVKI
 251 PFVPQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFWHI YPDENVHLD
 301 KLHAFWDKAY GFYTPETASV HRLSDDLNG GEALSATQRL ECGVL*

15 Further sequencing revealed the following DNA sequence <SEQ ID 195>:

20
 1 ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTTTGCA AGGTCATCGA
 51 CAATTTCGGC GACATCGGCG TTTCTGTGGC GCTCGCCCGT GTTTTGACCG
 101 GCGAACTCGG TTGGCAGGTG CATTGTGTGA CGGACGACGT GTCCGCGCTTG
 151 CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT
 201 TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
 251 TTCCCGATGC CGTTATCGAA ACTTTTGCTT GCGACCTGCC CGAAAATGTG
 301 CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGGAATATTT
 351 GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGCAGG
 401 AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
 25 451 GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATAACCGA
 501 AGCCCTGCGC CGGCGGCTGG TGCTGCCCCG AAAAAACGCC CCCGAATGGC
 551 TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
 601 CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
 651 CGACAGCCTC AAACAAAGCG GCGTTATTCC GCAAAACGCC CTGCAAAATg
 30 701 aaggcgGTGT CTTTCagacG gcatccgTcC gccttGTCAA AAtcCCGTTC
 751 GTGCGGCAAC AGGAcTTCGA CAAATTGCTG CAcctcgCG ACTGCGCCGT
 801 GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCTC
 851 TTTTTTGGCA CATCTACCG CAAGACGAGA ATGTCCATCT CGACAAATC
 901 CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCG AAACCGCATC
 35 951 GGTGCAACGC CTCCTTTCGG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
 1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAACGGC
 1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC
 1101 CGCATCCGAA AAACCTCGCG CTTTGTGTTT AAAGCATCAA AAAATACGCT
 1151 AG

40 This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

45
 1 MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDVVSAL
 51 RALCPDLPDV PFVHQDIHVR TWHSDAADID TAPVDPDAVIE TFACDLPENV
 101 LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYF WFMGFSEKSG
 151 GLIRERDYRE AVRFDTEALR RRLVLP EKNA PEWLLFGYRG DVWAKWLDMW
 201 QQAGSLMTLL LAGAQIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
 251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDKL
 301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLC WQTLQQHQNG
 351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

50
 orf32-1.pep 10 20 30 40 50 59
 MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV
 orf32ng-1 MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV
 55
 orf32-1.pep 60 70 80 90 100 110 119
 PCVHQDIHVRTWHSDAADIDTAPVDPDVVETFACDLPENVLNIIRRHKPLWLNWEYLSAE
 orf32ng-1 PFVHQDIHVRTWHSDAADIDTAPVDPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE
 60
 120 130 140 150 160 170 179

-155-

	orf32-1.pep	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNA
	orf32ng-1	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYREAVRFDTEALRRRLVLPEKNA
5		130 140 150 160 170 180
	orf32-1.pep	180 190 200 210 220 230 239
	orf32ng-1	SEWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSILKQSGVIPQDALQNDGDFVQT
10		190 200 210 220 230 240
	orf32-1.pep	240 250 260 270 280 290 299
	orf32ng-1	ASVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFFFWHIYPQDENVHLDKL
15		250 260 270 280 290 300
	orf32-1.pep	300 310 320 330 340 350 359
	orf32ng-1	HAFWDKAHGFYTPETVSAHRRLLSDDLNGGEALSATORLECWQTLQOHQNGWRQGAEDWSR
20		310 320 330 340 350 360
	orf32-1.pep	360 370 380
	orf32ng-1	YLFQGPSAPEKLA AFVSKHQKIRX
25		370 380

30 On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 35 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 24

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

45

```

1   ..TTGTTCTGTC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG
51  GTTTCGGGNC AAAGACCTG TAAATCAGGC GGTGTGCGG CTGTATNCGG
101 ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
151 AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTCGG TATTGTTGCT
201 GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAGCAGC CTGTTGAGCA
251 ATGCCGCTTC GGTACGCGCG GTGGAATGT TGGCATGGCT GCCGTCGAAA
301 CTCGGTTTCC CTGTCCCCGA TCGCGGTCG GTCATCGAAG GCCGTCTGAA
351 CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
401 TCGCCTGCTA NGGCATCCTG CCGCGCCTG..

```

50 This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

```

1   ..LFLRVKVGRF FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
51  SLWLCTLLGM LVSVLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK
101 LGFPVPDARS VIEGRLNGNI ADARAWSGLL VXSIACXGIL PRL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```

      1 ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
      51 AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
    101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGGAGATG
      151 ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCCTGCGGG
    201 GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTCGATTT TTTACCGGTT
      251 TTTCAGTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
    301 GTTTTGGCGG GCGTGTTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
      351 GGCAATGTTG TTCCTGCGTG TGAAGTGGG GCGTTTTTTC AGCAGTCCGG
    401 CGACGTGGTT TCGGGGCAAA GACCTGTAA ATCAGGCGGT GTTGCGGCTG
      451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGGAATA TAGGCGCAAC
    501 GTCGCACAGC CTGTGGCTCT GCACGTGCTC CGGAATGCTG GTGTGCGTAT
      551 TGTGCTGCTT TTTGGTGGG CAATATACGT TCAACTGGGA AAGCAGCTG
    601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
      651 GTCGAACTC GGTTCCTG TCCCGATGC GCGGGCGGTC ATCGAAGGCC
    701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTGGGG GCTGCTGGTC
      751 GGCAGTATCG CCTGCTACGG CATCTGCCG CGCTGCTGG CTTGGGTAGT
    801 GTGTAAAATC CTTTTGAAAA CAAGCGAAAA CGGATTGGAT TTGAAAAAGC
      851 CCTATTATCA GCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
    901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAAA TCATCTTGAA
      951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
   1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
   1051 ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
   1101 GCGGCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCG GACCGCGGCG
   1151 TGTTGCGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CCGCGCGGTG
   1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
   1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
   1301 CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

```

    30      1 MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
      51 IDNRMLRET LERVAGSFW LWVVAATFAF FTGFSVTYLL MDNQGLNFFL
   101 VLAGVLGMNT LMLAVWLAML FLRVKVGRRF SSPATWFRGK DPNQAVLRL
   151 YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
   201 LSNAASVRV EMLAWLPSKL GFPPVDARAV IEGRLNGNIA DARAWSGLLV
   251 GSIACYGILP RLLAWVCKI LLKTSENGLD LEKPYYQAVI RRWNKITDA
   301 DTRRETVS AV SPKII LNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
   351 TNREQVAAL TELKQKPAQL LIGVRAQTPV DRGVLRQIVR LSEAAQGGA V
   401 VQLLAEQGLS DDLSEKLEHW RNALAECCAA WLEPDRAAQE GRLKQD*

```

Computer analysis of this amino acid sequence gave the following results:

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N. meningitidis*:

```

    45      orf33.pep      LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR
      orf33a      LMDNQGLNFFLVLAGVXGMNTLMLAVWLAMLFLRVKVGRRFFSSPATWFRGKDPVNQAVLR
      90      100      110      120      130      140

    50      orf33.pep      LYXDEWRXTSVRWKIXATSHSLWLC TLLGMLVSVLLLLLV RQYTFNWESTLLSNAASVRA
      orf33a      LYADEWRXPSVRWKIGATSHSLWLC TLLGMLVSVLLLLLV RQYTFNWESTLLGDSVVRL
      150      160      170      180      190      200

    55      orf33.pep      VEMLAWLPSKLGFPVPDARSVIEGR LNGNIADARAWSGLLVXS IACXGILPRL
      orf33a      VEMLAWLPAKLGFPVPDARAVIEGR LNGNIADARAWSGLLVGSI ACYGILPRL LAWAVCK
      210      220      230      240      250      260

    60      orf33a      ILXXTSENGLDLEKXXXXXIRRWQNKITDADTRRETVS AVSPKIVLNDAPKWAVMLETE
      270      280      290      300      310      320

```

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

```

1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGAAGA
51 AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC
101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGAAGATG
5   151 ATCGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TCGGTGCGGG
201 GTCGTTCTGG TTGTGGGTGG CGGCGGCGAC GTTTCGTTTT NTTACCGNTT
251 TTTTCAGTTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTTG
301 GTTTTGCGCG GCGTGNITGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTCCTGCGCG TGAAAGTGGG GCGTTTTTTC AGCAGTCCCG
10  401 CGACGTGGTT TCGGGGCAAA GACCCTGTCA ATCAGGCGGT GTTGGCGCTG
451 TATGCGGACG AGTGGCGGCN ACCTTCGGTA CGTTGAAAAA TAGGCGCAAC
501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 TGTGCTGCT TTTGGTGGCG CAATATACGT TCAACTGGGA AAGCACGCTG
601 TTGGGCGATT CGTCTTCGGT ACGGCTGGTG GAAATGTTGG CATGGCTGCC
15  651 TGCGAACTG GGTTCCTCCG TGCCTGATGC GCGGGCGGTC ATCGAAGGTC
701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCGGG GCTGCTGGTC
751 GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTCTTGG CTTGGGCGGT
801 ATGCAAAATC CTTNTGNAAA CAAGCGAAAA CGGCTGGAT TTGAAAAAGC
851 NNNNNNTCN NNCGNTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
20  901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCCGAAAA TCGTCTTGAA
951 CGATGCGCGG AAATGGCGGG TCATGCTGGA GACCGAATGG CAGGACGCGC
1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGGCAACTG CTTATCGGCG TCGCGCCCCA AACTGTGCCC GACGCGCGCG
25  1151 TGTTCGCGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC
1301 CCGACAGAGC GGGCAGGAA GGCCTCTGA AAACCAACGA CCGCACTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 202>:

```

30  1  MLNPSRKLVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
51  51  IDNRNMLRET LERVAGSEFW LWVAAATFAE XTXFSVTYLL MDNQGLNFFL
101 101  VLAGVXGMNT LMLAVWLAML FLRVKVGREF SSPATWFRGK DPNVQAVLRL
151 151  YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201 201  LGDSSSVRLV EMLAWLPAKL GFFVPDARAV IEGRNLGNIA DARAWSGLLV
35  251  GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
301 301  DTRRETSAV SPKIVLNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351 351  ANREQVAAL TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401 401  VXLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRKLTNDRT*

```

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

```

40  10      20      30      40      50      60
    orf33a.pep  MLNPSRKLVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAKMIDNRNMLRET
    orf33-1     MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRNMLRET
45  10      20      30      40      50      60
    orf33a.pep  LERVAGSEFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
    orf33-1     LERVAGSEFWLVVAAATFAFFTGFSTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
50  70      80      90      100     110     120
    orf33a.pep  LERVAGSEFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
    orf33-1     LERVAGSEFWLVVAAATFAFFTGFSTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
55  70      80      90      100     110     120
    orf33a.pep  FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
    orf33-1     FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
60  130     140     150     160     170     180
    orf33a.pep  FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
    orf33-1     FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
65  130     140     150     160     170     180
    orf33a.pep  FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
    orf33-1     FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
60  190     200     210     220     230     240
    orf33a.pep  VSVLLLLLVRQYTFNWESTLLGDSSSVRLVEMLAWLPAKLGFPVPDARAVIEGRNLGNIA
    orf33-1     VSVLLLLLVRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
65  190     200     210     220     230     240
    orf33a.pep  VSVLLLLLVRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
    orf33-1     VSVLLLLLVRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
65  250     260     270     280     290     300
    orf33a.pep  DARAWSGLLVGSIACYGILPRLAWAVCKILXXTSENGLDLEKXXXXXXIIRRWQNKITDA
    orf33-1     DARAWSGLLVGSIACYGILPRLAWAVCKILXXTSENGLDLEKXXXXXXIIRRWQNKITDA

```

-158-

5	orf33-1	DARAWSGLLVGS	250	260	270	280	290	300
	orf33a.pep	DTRRET	310	320	330	340	350	360
10	orf33-1	SAVSPKIVLNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGV	310	320	330	340	350	360
	orf33a.pep	SAVSPKII	370	380	390	400	410	420
15	orf33-1	LDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVATNREQVA	310	320	330	340	350	360
	orf33a.pep	LDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVATNREQVA	370	380	390	400	410	420
20	orf33-1	ALTECGAAWLEPDRAAQEGRLKTNDR	430	440	450			
	orf33a.pep	ALTECGAAWLEPDRAAQEGRLKTNDR	430	440	450			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N. gonorrhoeae*:

25	orf33.pep	LFLRVKVGRFFSSPATWFRXKDPVNQAVLR	30
	orf33ng	LMDNQGLNFFLVLAGVLGMNTLMLAVWLATLFLRVKVGRFFSSPATWFRGKGPVNQAVLR	100
30	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA	90
	orf33ng	LYADQWRQPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA	160
35	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGRNLNGNIADARAWSGLLVXSIA	143
	orf33ng	VEMLAWLPSKLGFPVPDARAVIEGRNLNGNIADARAWSGLLVGSIVCYGILPRLLAWVVCK	220

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

40	1	MIDRDRMLRD	TLERVRAGSF	WLWVVVASMM	FTAGFSGTYL	LMDNQGLNFF
	51	LVLAVGLGMN	TLMLAVWLAT	LFLRVKVGRF	FSSPATWFRG	KGPVNQAVLR
45	101	LYADQWRQPS	VRWKIGATAH	SLWLCTLLGM	LVSLLLLLV	RQYTFNWEST
	151	LLSNAASVRA	VEMLAWLPSK	LGFPVPDARA	VIEGRNLNGNI	ADARAWSGLL
50	201	VGSIVCYGIL	PRLLAWVVCK	ILLKTSENGL	DLEKTYQAV	IRRWQNKITD
	251	ADTRRET	VSA VSPKIVLND	PKWALMLETE	WQDQGWFEGR	LAQEWLDKGV
55	301	AANREQVAAL	ETELKQKPAQ	LLIGVRAQTV	PDRGVLRQIV	RLSEAAQGGG
	351	VVQLLAEQGL	SDDLSEKLEH	WRNALTECGA	AWLEPDRVAQ	EGRLKDQ*

Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

50	1	ATGTTGaatC	CATCCCgaAA	ACTGgttgag	ctGgTCCgtA	Ttttgaataa
	51	agggggtTTT	attttcagcg	gcgatacctgt	gcaggcgacg	gaggctttgc
55	101	gccgcgtgga	cggcAGTACG	GAggAaaaaa	tcttccgtcg	GGCGGAGAtg
	151	atcgACAGGg	accgtatgtt	gcgggACaCg	TtggaacGTG	TGCGTGcggg
60	201	gtcgtTctgG	TTATGGGTGG	TggtggCatC	gATGATGTtt	aCCGCCGGAT
	251	TTTCAGgcac	ttatCttCTG	ATGGACaatC	AGGGGcTgAA	TtTCTTTTA
65	301	GTTTTTggcgG	GAGTGTtggG	CATGaatacG	ctgATGCTGG	CAGTATGgtt
	351	gGCAACGTG	TTCTGCGCG	TGAAAGTGGG	ACGGTTTTTC	AGCAGTCCGG
70	401	CGACGTGGTT	TCGGGGCAAA	GGCCCTGTAA	ATCAGGCGGT	GTTGCGGCTG
	451	TATGCGGACC	AGTGGCGGCA	ACCTTCGGTA	CGATGGAAAA	TAGGCGCAAC
75	501	GGCGCACAGC	TTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTCGGTAT
	551	TGCTGCTGCT	TTTGGTGGCG	CAATATACGT	TCAACTGGGA	AAGCACGCTG
80	601	TTGAGCAATG	CGCTTCGGT	ACGCGCGGTG	GAAATGTTGG	CATGGCTGCC
	651	GTCGAAACTC	GGTTTCCCTG	TCCCGATGC	GCGGGCGGTC	ATCGAAGGTC
85	701	GTCTGAACGG	CAATATTGCC	GATGCGCGGG	CTTGGTCGGG	GCTGCTGGTC
	751	GGCAGTATCG	TCTGCTACGG	CATCTGCCG	CGCCTCTTGG	CTTGGGTAGT

-159-

801 GTGTAAATC CTTTGA AAA CAAGCGAAAA CGGattgGAT TTGGAAAAAA
 851 CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
 901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCgaAAA TCGTCTTGAA
 951 CGATGCGCCG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC
 1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
 1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
 1101 GGCGCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG
 1151 TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
 1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC
 1301 CTGACAGGGT GGGCAGGAA GGCCGTTTGA AAGACCAATA A

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

1 MLNPSRKLVE LVRILNKGGF IFSGDPVQAT EALRRVDGST EEKIFRRAEM
 51 IDRDRMLRDT LERVAGSFW LWVVVASMMF TAGFSGYLL MDNQGLNFFL
 101 VLAGVLGMNT LMLAVWLATL FLRVKVGRRF SSPATWFRGK GPNQAVLRL
 151 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
 201 LSNAASVRV EMLAWLPSKL GFPVPDARAV IEGRNLGNIA DARAWSGLLV
 251 GSIVCYGILP RLLAWVVCKI LLKTSENGLD LEKTYQAVI RRWQNKITDA
 301 DTRRETSAV SPKIVLNDAP KWALMLETW QDGQWFEGRL AQEWLDKGVA
 351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVLQIVR LSEAAQGGAV
 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ*

ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

		10	20	30	40	50	60
25	orf33-1.pep	MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRMLRET					
	orf33ng-1	MLNPSRKLVELVRILNKGGFIFSGDPVQATEALRRVDGSTEEKIFRRAEMIDRDRMLRDT					
		10	20	30	40	50	60
30	orf33-1.pep	LERVRAGSFWLWVVAATFAFFTGFSVTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLAML					
	orf33ng-1	LERVRAGSFWLWVVASMMFTAGFSGYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLATL					
		70	80	90	100	110	120
35	orf33-1.pep	FLRVKVGRRFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML					
	orf33ng-1	FLRVKVGRRFSSPATWFRGKGPVNQAVLRLYADQWRQPSVRWKIGATAHSLWLCTLLGML					
		130	140	150	160	170	180
40	orf33-1.pep	VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
	orf33ng-1	VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
		190	200	210	220	230	240
45	orf33-1.pep	DARAWSGLLVGSIACYGILPRLAWVVCKILLKTSENGLDLEKPYQAVIRRWQNKITDA					
	orf33ng-1	DARAWSGLLVGSIACYGILPRLAWVVCKILLKTSENGLDLEKTYQAVIRRWQNKITDA					
		250	260	270	280	290	300
50	orf33-1.pep	DTRRETSAVSPKII LNDAPKWAVMLETWQDGQWFEGRLAQEWLDKGVA TNREQVAALE					
	orf33ng-1	DTRRETSAVSPKIVLNDAPKWALMLETWQDGQWFEGRLAQEWLDKGVA ANREQVAALE					
		310	320	330	340	350	360
55	orf33-1.pep	TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
	orf33ng-1	TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
		370	380	390	400	410	420
60	orf33-1.pep	RNA LAECGA AWLEPDRAAQEGRLKDQX					
	orf33ng-1	RNA LAECGA AWLEPDRAAQEGRLKDQX					
		430	440				
65	orf33-1.pep						

orf33ng-1 RNALTECGAAWLEPDRVAQEGRLKDQX
430 440

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is
5 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```

10      1  ..CAGAAGAGTT TGTGAGAAT TTCTTTATGG GTTTGGGCG GCGTGTTTTT
      51  CCGGGGTGCC GGTCTGGTAT GGTTCCTTTT GGGCGTTTCT TT.GAGTGCG
      101  CCTGTTTTTC GGGTGTCTTCT TTTCGGGGTT CGGGACGGGG GACGTTTGTG
      151  GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTG TTCC.GGCGT
      201  CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTTG CG..GTTTGA
      251  CCCGGTTTTT CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
      15  301  TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
      351  GGGTTGGGCG GCATCTTGT CCGACTACGC CGTTTGGCAG CCAGAATTCTG
      401  GTTTCGCGGG GGCTGTCGGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
      451  GTCC..

```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```

20      1  ..QKSLSRISLW GLGGVFFGVS GLVWFSLGVS XECACFSGVS FRGSGRGTFV
      51  GSTGVSLSVF SACVXGVVRL PVGLSCVGRL XXLTRFFLGA AGDVILLPLS
      101  SVPSGCAGSD EAAWCSGWA ASCPTTFPGS QNSVSRGLSV CCGSA*RVLS
      151  S..

```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

```

25      1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
      51  GCCGGGTCAG AATAGGTTGT CCAGAATTTT TTTATGGGGT TTGGGCGGCG
      101  TGTTTTTTCG GGTGTCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
      151  GGCTGCGCCT GTTTTTCGGG TGTTCCTTTT CGGGGTTCGG GACGGGGGAC
      201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTTCA GCTTGTGTTC
      30  251  CGGCGTCGTC CGGCTCGCTG TCGGTTTGAG CTGTGTCGGC AGGTTGCCGT
      301  TTGACCCGGT TTTTCTTGGG TCGGCGAGGG GACGGCAGTC CGCTGCCGCT
      351  TTCGTCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT
      401  GTTCGGGTTG GCGGCGCATCT TGTCCGACTA CGCCGTTTGG CAGCCAGAAT
      451  TCGGTTTCGC GGGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT
      35  501  GTCGCGTTC GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
      551  TGGCGGCGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
      601  AGCCTGAAGG GTTTGTTCGG TTTTTTTGCC ATTTTGATTG TGCTTTTGGG
      651  GTGTCGGGCA ATGCCGTCG AAGGCGGTTT AGACGGCATT GCCGAGTCAG
      701  CGTTGGACGT AGTTTGGTA GAGGGTGATG ACTTTTGTGA CGCCGACGCT
      40  751  GGTGCTGACT TTTTGGGTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC
      801  CCATAACGTA GGTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
      851  GTGGCGGGGC TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTTCCTA
      901  AGTGTGCGCG GCGATGTCGC CGGCAGTGCG CGGCAGGGAG GCGACGGTAA
      951  TATAGTTGTA CACGCCTTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG
      45  1001  ACGAACTGTT TTTCGCCTTC GGTGGCGACT TGTCCGAGCA GCAGCAGGTG
      1051  GCGGTTGTAG CCGACGACGG AGATTGTTGGG CGTGTAGCCT TTGGTTTGGT
      1101  TGTTTTGGCG CAGATAGGAA CGGGCGGTGG TTTCGATACG CAACGCCATA
      1151  ACGTTGTCGT CGGTTTGGCG GCCGGTGGTT CGGCGGTCTGA CGGCGGATTT
      1201  CGCGCCGACG GCGGCGCTTC CGATTACTGC GCTGACGCAG CCGCTAAGGG
      50  1251  CAAGGCTGAA AATGGCGGCA ATCAGGGTGC GGACGGTGTG CGGTTTGGGT
      1301  TTCATCGGGT GCTTCCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC
      1351  CATGCCGTCT GA

```

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

```

55      1  MMMPFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLGVS
      51  GCACFSGVSF RSGRGRTFVG STGVSLSVFS ACVPASSGCL SV*AVSAGCG
      101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTFPGSQN
      151  SVSRGLSVCC GSA*RVLSPF GLNVLTMPIA NAPMAAIQMS NTARIRSLGV

```

5

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 *meningitidis:*

15

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

50

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

Homology with a predicted ORF from *N.gonorrhoeae*

orf34.pep QKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE-----CAC 35

-163-

```

    orf34ng      MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC      60
    orf34.pep    FSGVSFRGSGRGTFVGGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXXLTRFFLGA      90
5    orf34ng      FSGVSFRGSGWGAFVGGSTGVSLSVFSACVP-----VPVNESAARAASEGR--GLTRFFLGA      114
    orf34.pep    AGDVILLPLSSVPSGCAGSDEAAWWCSGWAASCPTTFPGSQNSVSRGLSVCCGSAXRVLS      150
10   orf34ng      AGDGSPLPLSSVPSGCAGSDEAAWWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLS      174
    orf34.pep    S                                                                    175
    orf34ng      PFGLNVLTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD      234

```

15 The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

```

1   ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
51  GCCGGGTCAA AAGAGGTTGT CGAGAATCTC TTTATGGGGT TTGGCCGGCG
101 TGTTCCTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTT
151 TCTTTGGGTG TTTCTTTGGG CTGCGCCTGT TTTTCGGGTG TTTCTTTTCG
20  GGGTTCGGGA TGGGGGGCGT TTGTGGGCAG TACGGGGGTT TCTTTGAGTG
251 TGTTCCTCAGC TTGTGTCCG GTGCCGGTTA ACGAATCGGC TGCCCCGGCC
301 GCATCCGAAG GCGCGGTTT gACCCGGTTT TTCTTGGGTG CCGCAGGGGA
351 CCGCAGTCCG CTGCCGCTTT CTTCTGTGCC GTCCGGCTGT GCGGGTTCGG
401 ATGAGGCGGC GTGGTGGTGT TCGGGTTGGG CGGCATCTTG TCCGACGGCG
25  CCGTTTGCCA GCCAGAAATC GGTTCGCGG GGGCTGTGCG TGTGTTGCGG
501 TTCGGTTTGG AGGGTTTGTG CGCCGTTCCG GTTGAATGTG CTGACGATGC
551 CTAAGTCCAA TCGCGCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
601 ATCAGGAGTT TGGGGGTGAG CCTGAAGGGT TTGTTCGGTT TTTTGGCCAT
651 TTTGATTGTG CTTTTGGGGT GTCGGGCAAT GCCGTCTGAA GCGGGTTCAG
30  ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTTGGTAGA GGGTAATGAC
751 TTTTGTGACG CCGAcggTGG TGCTGACTTT TTGGGTAATC TCGCCTGTT
801 CTTTCGGGGT GAGGATGCCC ATAACGTAGG TTACATTGCC GTAGGTAATG
851 ATTTTGACGC GCGCCTGTGT AGCGGGGCTG ATGCCAGCA GcgtgCGCG
901 GACTTTGACG GTGTTCCAAG TGTCGCGCGC GATGTCGCCC GCAGTGCAGC
35  GCAGGGAGGC GACGGTAATG TAGTTGTATA CGCCTTCGGC GGCCTGTTTC
1001 GAACGTGCAA TCTGACCGAC GAACTGTTT TCGCCTTCGG TGGCGACTTG
1051 TCCGAGCAGC AGCAGGTGGC GGTGTAGACC GACGACGGAG ATTTGGGGCG
1101 TGAGCCTTT GGTGTGGTGT TTTTGGCGCA GGTAGGAACG GCGGTGGTT
1151 TCGATACGCA ACGCCATAAC GTTgtCATCG GTTtgcgcgc CGGTGGTTcg
40  gCGGTCGATG ACGGATTTTG CGCCGACGGC GGCCCGCCCG ACGACTGCGC
1251 TGAAGCAGCC GCCGAGGGCA AGGCTGAGGA CGCGGCAAT CAGGGTCCGG
1301 ACGGTGTGTG GTTTGGGTTT CATCGGGGAC TTCCTTTCTT GGGCGTTTCA
1351 GACGGCATTG CTTTGCGCCA TGCCGTCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 214>:

```

45  1   MMMPFIMLPW IAGVPAVPGQ KRLSRISLWG LAGVFFGVSG LVWFSLVGSF
    51  SLGVSLGCAC FSGVSFRGSG WGAFFVGSTGV SLSVFSACVP VPVNESAARA
101  ASEGRGLTRF FLGAAGDGSP LPLSSVPSGC AGSDEAAWWC SGWAASCPTA
151  PFGSQNSVSR GLSVCCGSVW RVLSPFGLNV LTMPTANAPM AVIQMSNTAR
201  IRS LGVSLKG LFGFFAILIV LLGCRAMPSE GSDGIAESA LDVVLVEGND
50  251  FLYADGGADF LGNLR LFFGG EDHNVGYIA VGNDFDARLC SGADAQORGA
    301  DFRVPSVAG DVAR SARQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL
    351  SEQQQVAVVA DDGDLGRVAF GLVVLAQVGT GGGFDTQRHN VVIGLRAGGS
    401  AVDDGFCADG GPADDCAEAA AEGKAEDGGN QGADGVWFGE HRGLPFLGVS
    451  DGIALRHAV*

```

55 ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

```

    orf34-1.pep    MMMPFIMLPWIAGVPAVPGQNRSLSRISLWGLGGVFFGVSGLVWFSLVGS-----LGCAC
60   orf34ng      MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC
    10           20           30           40           50
    orf34-1.pep    FSGVSFRGSGRGTFVGGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
65   orf34ng      FSGVSFRGSGWGAFVGGSTGVSLSVFSACVPVPVNESAARAASEGRGLTRFFLGAAGDGSP
    60           70           80           90          100          110

```

-164-

		70	80	90	100	110	120
		120	130	140	150	160	170
5	orf34-1.pep	LPLSSVPSGCAGSDEAAW	CSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV				
	orf34ng	LPLSSVPSGCAGSDEAAW	CSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLSPFGLNV				
		130	140	150	160	170	180
10	orf34-1.pep	180	190	200	210	220	230
	orf34ng	LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
		190	200	210	220	230	240
15	orf34-1.pep	240	250	260	270	280	290
	orf34ng	LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
20		250	260	270	280	290	300
	orf34-1.pep	300	310	320	330	340	350
	orf34ng	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGFLFGTCNLTDelfFAFGDDLSEQQQVAVVA					
25		310	320	330	340	350	360
	orf34-1.pep	360	370	380	390	400	410
	orf34ng	DDGDLGRVAFGLVVLAAQIGTGGGFDTQRHNVVVLGRAGGSVDGGFRADGGASDYCADAA					
30		370	380	390	400	410	420
	orf34-1.pep	420	430	440	450		
	orf34ng	AKGKAENGNGQADGVRFGFHRVLPFLGVSDGIALRHAVX					
35		430	440	450	460		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```

45      1  ATGAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
     101  CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
     151  CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAATCCAA GCCGAGCTGG
     201  AGAAAAAAGG CTACACCGTC AACTGGTTCG AGTTTACCGA CTATGTACGC
     251  CCGAATCTGG CATTGGCTGA GGCAGATTG

```

50 This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```

      1  MKTFFKTLA AALALILAA G.QKDSAPAA SASAAADNGA AKKEIVFGTT
     51  VGDGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```

55      1  ATGAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GGCGGTCAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
     101  CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
     151  GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
     201  GAAAAAAGGC TACACCGTCA AACTGGTTCG GTTTACCGAC TATGTACGCC

```

-165-

251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC
 301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATACCGA
 351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
 401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
 451 CCGTCCAACT TCGCCGCGT CTTGGTGATG CTCGACGAAC TGGGTGGAT
 501 CAAACTCAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG
 551 CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
 601 CCGCGTAGCC GCGCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAAATGAAG
 851 GCGCAGCCAA ATAA

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

15 1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AKKEIVFGTT
 51 VGDGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
 101 KPYLDDEFKE HNLDTVEVFQ VPTAPLGLYP GKLSLEEVK DGSTVSAPND
 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
 201 PRSRADVDFV VVNGNYAISS GMKLTEALFQ EPSFAYVNS AVKTADKDSQ
 20 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis*:

25 orf4.pep 10 20 30 40 50 59
 MKTFFKTLA AALALILAACG-QKDSAPAA SASAADNGA AKKEIVFGTT VGDGDMVKE
 orf4a MKTFFKTLA AALALILAACG GQKDSAPAA SASAADNGA AXKEIVFGTT VGDGDMVKE
 30 60 70 80 90
 orf4.pep QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL
 orf4a XIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDIN VQHXXYLDDXKXHNLDITXVXQ
 35 70 80 90 100 110 120
 orf4a VPTAPLGLYP GKLSLXVXK GSTVSAPND PXXFXRVLV MLDELGXIKLKD XIXXXXXX
 130 140 150 160 170 180

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

40 1 ATGAAAACCT TCTTCAAAC CCTTCCGCC GCGCACTCG CGCTCATCCT
 51 CGCCGCCGTC GCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
 101 CCGCCGCCGA CAACGGCGCG GCGAANAAG AAATCGTCTT CGGCACGACC
 151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA
 201 GAAAAAAGGC TACACCGTCA AACTGGTCTGA GTNTACCGAC TATGTGCGCN
 45 251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTNCAACAC
 301 ANACNCTATC TTGACGACTN CAAAAANAA CACAATCTGG ACATACCNN
 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
 401 AATCGCTGGA NNAAGTCAA GANGGCAGCA CCGTATCCGC GCCCAACGAC
 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTNGAT
 50 501 CAAACTCAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANN
 551 NNNNANNNT NNNNNNNNN NNNNNCNCG NNNNNNNANN NNNNNNNNN
 601 NCGNNTNNNN NNGCNNNNNT NNANNTNNN NNCNNCNNN NNNNTNNNN
 651 NANNANNAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
 55 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAAATGAAG
 851 GCGCAGCCAA ATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AXKEIVFGTT

-166-

51 VGDFGDMVKE XIQPELEKKG YTVKLVEFTD YVRXNLALAE GELDINVXQH
 101 XXYLDDXKKX HNLDITXVXQ VPTAPLGLYP GKLKSLXXVK XGSTVSAPND
 151 PXXFXRVLM LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX
 201 XXXXAXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
 5 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

1 ATGAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
 51 CGCCGCTGC GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
 10 101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCAGCACC
 151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
 201 GAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTGCGCC
 251 CGAATCTGGC ATTGGCTGAG GCGAGTTGG ACATCAACGT CTCCAACAC
 301 AAACCCTATC TTGACGACTT CAAAAAGAA CACATCTGG ACATCACC
 15 351 AGTCTTCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
 401 AATCGCTGGA AGAAGTCAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
 451 CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGTTGGAT
 501 CAAACTCAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
 551 CCGAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
 20 601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
 25 851 GCGCAGCCAA ATAA

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
 51 VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVXQH
 101 KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLKSLXEVK DGSTVSAPND
 30 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKN IVELEAAQL
 201 PRSRADVFA VVNGNYAISH GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

35	orf4a-1	10	20	30	40	50	60
		MKTFFKTLA AALALILAACGGQKDSAPAA SASAAADNGA AKKEIVFGTTVGDFGDMVKE					
	orf4-1	MKTFFKTLA AALALILAACGGQKDSAPAA SASAAADNGA AKKEIVFGTTVGDFGDMVKE					
40	orf4a-1	70	80	90	100	110	120
		QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVXQH KPYLDDFKKE HNLDITEVFQ					
	orf4-1	QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVXQH KPYLDDFKKE HNLDITEVFQ					
45	orf4a-1	130	140	150	160	170	180
		VPTAPLGLYPGKLKSLXEVKDGSTVSAPNDPSNFARVLV LDELGWIKLK DGINPLTASK					
	orf4-1	VPTAPLGLYPGKLKSLXEVKDGSTVSAPNDPSNFARVLV LDELGWIKLK DGINPLTASK					
50	orf4a-1	190	200	210	220	230	240
		ADIAENLKN IVELEAAQL PRSRADVFA VVNGNYA ISSGMKLTEALFQ EPSFAYVNWS					
55	orf4-1	ADIAENLKN IVELEAAQL PRSRADVFA VVNGNYA ISSGMKLTEALFQ EPSFAYVNWS					
60	orf4a-1	250	260	270	280		
		AVKTADKDSQWLKDVTEAYNSDAFKAYAHK RFEGYKSPAA WNEGAAKX					
	orf4-1	AVKTADKDSQWLKDVTEAYNSDAFKAYAHK RFEGYKSPAA WNEGAAKX					

-167-

Homology with an outer membrane protein of *Pasteurella haemolítica* (accession q08869).

ORF4 and this outer membrane protein show 33% aa identity in 91aa overlap:

```

5      lip2.pasha      MNFKLLGVALVSALALTACKDEKAQAP----
      ORF4      VXTPNPDGRTPCPSFLFETATTSGENMKTFFKTLSTASLALILAACGGQKDSAPAA
                110      120      130      140      150

10     lip2.pasha      30      40      50      60      70      80
      ORF4      -ATTAKTENKAPLKVGMTGPEAQMTEVAVKIAKEKYGLDVELVQFTEYTQPNAAALHSKD
                : : : | : | : : : : : : : : : : : : : : : : : : : : : : :
      ORF4      LPPPTTARRKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGE
                160      170      180      190      200      210

15     lip2.pasha      90      100      110      120      130      140
      ORF4      LDANAFQTVPYLEQEVKDRGYKLAIIGNTLVWPIAAYSKKIKNISLKDGDATVAIPNNAS
                |
                L.....

```

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae*:

```

25     orf4nm.pep      10      20      30
      orf4ng      MKTFFKTLSTASLALILAACGGQKDSAPAA
                ||||| : : : : : : : : : : : : : : : :
                200      210      220      230      240      250

30     orf4nm.pep      40      50      60      70      80      89
      orf4ng      SASA-AADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA
                || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      orf4ng      SAAAPSADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA
                260      270      280      290      300      310

35     orf4nm.pep      90
      orf4ng      EGEL
                ||||
      orf4ng      EGELDINVFQHKPYLDDFKKEHNLDITEAFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPN
                320      330      340      350      360      370

```

40 The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

```

      1  MKTFFKTLST ASLALILAAC GGQKDSAPAA SAAAPSADNG AAKKEIVFGT
     51  TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ
    101  HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN
    151  DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ
    201  LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
    251  QWLKDVTEAY NSDAFKAYAH KRFEGYKYPA AWNEGAAK*

```

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

```

50     1  atgAAAACCT TCTTCAAAAC cctttccgcc gccgcaCTCG CGCTCATCCT
     51  CGCAGCCTGc ggCggtcaAA AAGACAGCGC GCCCgcagcc tctgcCGCCG
    101  CCCCTTCTGC CGATAACGgc gCgGCGAAAA AAGAAAtcgt ctTCGGCAGC
    151  Accgtgggcg acttcggcgA TAtggTCAAA GAACAAATCC AagcCGAgct
    201  gGAGAAAAAA GgctACACcg tCAAattggt cgaatttacc gactatgtGC
    251  gCCCGAATCT GGCATTGGCG GAGGGCGAGT TGGACATCAA CGTCTTCCAA
    301  CACAAACCTT ATCTTGACGA TTCAAAAAA GAACACAACC TGGACATCAC
    351  CGAAGCCTTC CAAGTGCCGA CCGCGCCTTT GGGACTGTAT CCGGGCAAAC
    401  TGAATCGCT GGAAGAAGTC AAAGACGGCA GCACCGTATC CGCGCCCAac
    451  gACcgcTCCA ACTTCGCACG CGCCTTGGTG ATGCTGAACG AACTGGGTG
    501  GATCAAACTC AAAGACGGCA TCAATCCGCT GACCGCATCC AAAGCCGACA
    551  TCGCGGAAAA CCTGAAAAAC ATCAAAATCG TCGAGCTTGA AGCCGCACAA

```

-168-

5
 601 CTGCCGCGCA GCCGCGCCGA CGTGGATTTT GCCGTCGTCA ACGGCAACTA
 651 CGCCATAAGC AGCGGCATGA AGCTGACCGA AGCCCTGTTC CAAGAGCCGA
 701 GCTTTGCTTA TGTCAACTGG TCTGCCgtca AAACCGCCGA CAAAGACAGC
 751 CAATGGCTTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC
 801 CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG
 851 AAGGCGCAGC CAAATAA

This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:

10
 1 MKTFFKTLA AALALILAAC GGQKDSAPAA SAAAPSADNG AAKKEIVFGT
 51 TVGDFGDMVK EQIQAELEKK GTVTKLVEFT DYVRPNLALA EGELDINVFQ
 101 HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN
 151 DPSNFARALV MLNELGWIKL KGINPLTAS KADIAENLKN IKIVELEAAQ
 201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
 251 QWLKDVTEAY NSDAFKAYAH KRFEQYKYP AAWNEGAAK*

This shows 97.6% identity in 288 aa overlap with ORF4-1:

15
 orf4-1.pep 10 20 30 40 50 59
 MKTFFKTLA AALALILAACGGQKDSAPAA SAAAPSADNG AAKKEIVFGT TVGDFGDMVK
 orf4ng-1 MKTFFKTLA AALALILAACGGQKDSAPAA SAAAPSADNG AAKKEIVFGT TVGDFGDMVK
 20
 orf4-1.pep 60 70 80 90 100 110 119
 EQIQAELEKKGYTVTKLVEFTDYVRPNLALA EGELDINVFQHKPYLDDFKKEHNLDITEVF
 orf4ng-1 EQIQAELEKKGYTVTKLVEFTDYVRPNLALA EGELDINVFQHKPYLDDFKKEHNLDITEAF
 25
 orf4-1.pep 120 130 140 150 160 170 179
 QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTAS
 orf4ng-1 QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTAS
 30
 orf4-1.pep 180 190 200 210 220 230 239
 KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYA ISSGMKLTEALFQEPSFAYVNW
 orf4ng-1 KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYA ISSGMKLTEALFQEPSFAYVNW
 35
 orf4-1.pep 240 250 260 270 280
 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAXX
 orf4ng-1 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAXX
 40
 orf4-1.pep 240 250 260 270 280
 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAXX
 orf4ng-1 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAXX

45 In addition, ORF4ng-1 shows significant homology with an outer membrane protein from the database:

50
 ID LIP2_PASHA STANDARD; PRT; 276 AA.
 AC Q08869;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE 28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR. . . .
 SCORES Init1: 279 Initn: 416 Opt: 494
 Smith-Waterman score: 494; 36.0% identity in 275 aa overlap
 55
 orf4ng-1.pep 10 20 30 40 50
 MKTFFKTLA AAL--ALILAACGGQKDSAPAA SAAAPSADNG AAKKEIVFGT TVGDFGDM
 lip2_pasha MNFKLLGLVALVSALALTACKDEKAQAPATTA---KTENKAPLK---VGVMTGPEAQM
 60
 orf4ng-1.pep 60 70 80 90 100 110
 VKEQIQAELEKKGYTVTKLVEFTDYVRPNLALA EGELDINVFQHKPYLDDFKKEHNLDITE
 :: :: || | :||:|:|:|:|:| | :|| |:| | |:: |:: ::

-169-

	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTPQNAALHSKDLDANAFQTVPYLEQEVKDRGYKLAI	60	70	80	90	100	110
5	orf4ng-1.pep	AFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLT	120	130	140	150	160	170
	lip2_pasha	IGNTLVWPIAAYSKKIKNISELKDGATVAIPNNASNTARALLLQAHGLLKLKDPKN-VF	120	130	140	150	160	170
10	orf4ng-1.pep	ASKADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAIISSGMKLTE--ALFQEPSFA	180	190	200	210	220	230
	lip2_pasha	ATENDIENPKNIKIVQADTSLTRMLDDVELAVINNTYAGQAGLSPDKDGIIVESKDSP	180	190	200	210	220	230
15	orf4ng-1.pep	YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYPAAWNEGAAX	240	250	260	270	280	289
	lip2_pasha	YVNLVVSREDNKDDPRLQTFVKSFQTEEVFQEAALKLFNGGVVKGW	240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolítica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

	1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
	51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
40	101	GTTTATGCG	TGAACGACA	TTATTTCCAC	GGCAACCTCC	TCTTCTACCT
	151	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG	CTGGGCGGCG	GTCGGCAAAA
	201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT	GCTCATCGGC
	251	GACAACGGCA	GCGAATGGCT	CGACAGCGGA	CTCATGCGCG	CCATGAACGT
	301	CCTCATCGGC	GyGGCCATCG	CCATCGCCGC	CGCCAAACTG	CTGCCGCTGA
	351	AATCCACACT	GATGTGGCGT	TTCATGCTTG	CCGACAACCT	GGCCGACTGC
45	401	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC	AGGCGCATGA	CCCGCGAACG
	451	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA	AATCAACGCA	CGCATGGTCA
	501	AAAGCCGCAG	CCATCTCGCC	GCCACATCGG	GCGAAAGCTG	CATCAGCCCC
	551	GCCATGATGG	AAGCCATGCA	GCACGCCCAC	CGTAAATCG	TCAACACCAC
	601	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT	GCAATCTCCC	AAACTCAACG

5
 651 GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....
 701 GC AGACACGCCC GCCGCATCCG
 751 CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC
 801 ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA
 851 ATTCGCGCC TCGTCATCCT GCTGCAACGC ACCCGCCGA AATGGCTGGA
 901 TGCCACAGAA CGCCAACACC TGCGCCAAG CCTGCTTGA

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

10
 1PRRP RHAPVSRGDL LQGGGYARH GHRAGRGFGR FMAEPALFPR
 51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQRT
 101 HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
 151 AHDPRTPRGE HGENAPNORT HGQKPQPSRR HIGRKLHQPR HDGSHAARPP
 201 XNRQHHRAAP DHRQAAISQ TQRQNPAAAX PPLHTAPN..Q
 251 TRPPHPHRRH HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH
 301 PPQMAGCPRT PTPAPKPA*

15 Computer analysis of this amino acid sequence gave the following results:

Sequence motifs

ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N.gonorrhoeae*:

25
 orf8ng 1 MDRDDRLRRPRHAPVPRDLLQGGGYARYGHRAGRGFGRFMAEPALFPR 50
 orf8.pep 1PRRP RHAPVSRGDL LQGGGYARH GHRAGRGFGR FMAEPALFPR 44
 30
 orf8ng 51 QPPLLPDHRHGKRTGRLGGGRQKRLRPYVGGADDVHAHRRQRQRMARQRP 100
 orf8.pep 45 QPPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94
 35
 orf8ng 101 DARDERPHRRHRHRCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ 150
 orf8.pep 95 HARHERPHRRGHRHRRRQTA AAEIHTDVAF HACRQPGRMQ QNDCRNQQRQ 144
 40
 orf8ng 151 AYDARTFGAEYQONAPNORTHGQKPQPPRRHIGRKP HQPLHDGSHAARPP 200
 orf8.pep 145 AHDPRTPRGEHGENAPNORTHGQKPQPSRRHIGRKLHQPRHDGSHAARPP 194
 45
 orf8ng 201 QNRQHHRAAPDHRQAAISQTQRQNPAAAPPLHTAPNRPATNRRPHQRQ 250
 orf8.pep 195 XNRQHHRAAPDHRQAAISQTQRQNPAAAXPPLHTAPN.....Q 244
 50
 orf8ng 251 TRPPHPHRRHQPRTGSPRRTPPLPMAGFPLAQHQYASGNFRPRHPATH 300
 orf8.pep 245 TRPPHPHRRHQPRTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294
 orf8ng 301 PPQMAGCPRTPTPAPKPA* 319
 orf8.pep 295 PPQMAGCPRTPTPAPKPA* 313

50 The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

55
 1 MDRDDRLRRP RHAPVPRDL LQGGGYARY GHRAGRGFGR FMAEPALFPR
 51 QPPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP
 101 DARDERPHRR RHRHRCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
 151 AYDARTFGAE YQONAPNORT HGQKPQPPRR HIGRKP HQPL HDGSHAARPP


```

201 QNRQHHRAAP DHRROAISQ TORQRNPAAR PPLHTAPNRP ATNRRPHQRO
251 TRPPHPRHR HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH
301 PPQMAGCPRT PTPAPKPA*

```

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

```

10 1 ..GAAATCAGCC TGC GGTTCCGA CNACAGGCCG GTTTCGGTGN CGAAGCGGCG
51 GGATTTCGGAA CGTTTTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT
101 GGGCGTGGGT GGAACACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
151 CGCGATTTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
201 TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
251 TGCAGGAACA GTCGCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
15 301 GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
351 CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTACGCCGC AACGCCTGCG
401 TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
451 GGACATTATC TCGGAGA.GG AACCATCATG CCCGGTTTCC ACCTGATGAA
501 AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGTAAGC
20 551 GTTATCCTTT CCCGACCGG..

```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```

25 1 ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKAWVENG TFATVGSAPY
51 RDLSPFLGAEW AEKADGNVRI VGCAVCGEFK KAQVQEQLAR KIEWLPSSAQ
101 AGIRNHYRH PEEHGSDFWF NALGSRRFSR NACVVVSCGT AVTVDALTD
151 GHYLGXGTIM PGFHLKESL AVRTANLNRH AGKRYPFPT..

```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

```

30 1 ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
51 CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGCGAGCAGA TGCCGGCGCA CATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGTTGCGCC CATTGGCGGT
201 TTTCGATGCC GAAGCTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTTCAGA
251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
35 351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401 GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTTTGACCG GCCGCGATAT
451 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGTG GGCGCGCCTT
501 GTCGCGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTTGGTTG
551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
601 GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
40 651 GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
701 GCGGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAAACGCT GTGGTGGAA
751 CTGGACGCGG TGTTGTTGCA ATATGCGCGG GACGGATTTC CGCCTTTTGT
801 GCGGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTAAAGG CGTGGACGGA
45 901 CAAGGCGTTT TGCACCTTGA AACGGCAGAG GGCAACAGA CGGTGCTCAG
951 CCGCGAAATC AGCCTCGGT CCGACGACAG GCCGTTTCC GTGCCGAAGC
1001 GCGGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGGTGAAAAA CCGCACGTTT GCAACCGTCG GTAGCGCGCC
1101 GTACCGCGAT TTGTCGCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
50 1151 GAAATGTCGG CATCGTCGGT TGCGCTGTGT GCGGAGAATT CAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGCTGC CGTCTCCCG
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACACGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGCGAGC GCCGTTTCAG CCGCAACGCC
1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
55 1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501 CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTGCGCA GCGGCATGAT
1551 GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG

```

5	1	MTVLKLSHWR	VLAELADGLP	QHVSQALARMA	DMKPQQLNGF	WQQMPAHIRG
	51	LLRQHDGYWR	LVRPLAVFDA	EGLRELGERS	GFQTALKHEC	ASSNDEILEL
	101	ARIAPDKAHH	TICVTHLQSK	GRGRQGRKWS	HLRGLCELMFS	FGWVFDPRQY
	151	ELGSLSPVAA	INCFRRALSRL	GLDVQIKWPN	DLVVGRDKLG	GILIVTRTKG
10	201	GKTVAVVGIG	IVFRLPKEVE	NAASVQSLFQ	TASRRGNADA	AVLLETLLVE
	251	LDAVLLQYAR	DGFAPFVAEY	QAA NRDHGKA	VLLLRDGETV	FEGTVKGVVDG
	301	QGV LHLETA E	GKQTVVSGEI	SLRSDDRPVS	VPKRRDSERF	LLLDGGNSRL
	351	KWAWVENGTF	ATVGSAPYRD	LSPLGAEWAE	KADGNVIRVG	CAVCGEFKKA
15	401	QVQEQLARKI	EWLPSQAQAL	GIRNHYRHPE	EHGSDRWFNA	LGSRRFRSRA
	451	CVVVSCGTAV	TVDALTDGDH	YLGGTIMPGF	HLMKESLAVR	TANLNRRHAGK
	501	RYFPPTTTGN	AVASGMMDAV	CGSVMMMHGR	LKEKTGAGKV	VDVITITGGA
	551	AKVAEALPPA	FLAENTVRVA	DNLVIYGLLN	MIAAEGREYE	HI*

Homology with the baf protein of *B. pertussis* (accession number U12020).

```

orf61 23  LLLDGGNSRLKWAWVE-NGTFATVGSAPYR---DLSP LGAEWAEKADGNVRIVGCAVCG 77
      +L+D GNSRLK  W + +   A   AP      DL  LG  A      R +G  V  G
baf    3   ILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDL DALGRWLATLPRRPQRALGVNVAG 62

orf61 78  EFKKAQVQEQLAR---KIEWLPSSAQXGIRNHYRHPEEHGSDRW---FNALGSRRFRSRN 131
      +   +   L      I WL +   A G+RN YR+P++ G+DRW      L   +
baf    63  LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP 122

orf61 132 ACVVVSCGTAVTVDALTDGHHYLGXGTIMPGFHLMKESLAVRTANL 177
      +V S  GTA T+D +   D + G G I+PG  +M+ +LA  TA+L
baf    123 PLLVASFGTATTLDITIGPDNVFPG-GLILPGPAMMRGALAYGTAHL 167

```

ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of *N.*

```

                                10          20          30
orf61.pep      EISLRSDXRPVSVXKRRDSERFLLLDGGNS
                ||||| ||||| ||||| ||||| |||||
40 orf61a      TVFEGTVKGVDGQGV LHLETAEGKQT VVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNS
                290        300        310        320        330        340

                        40          50          60          70          80          90
orf61.pep      RLKWAWVENGT FATVGSA PYRDL SPLGA EWAEAK DGNVR I VGCAVCGEFKKA QVQEQLAR
                ||||| ||||| ||||| ||||| : ||||| ||||| ||||| ||||| ||||| |||||
45 orf61a      RLKWAWVENGT FATVGSA PYRDL SPLGA EWAEK VDGNVR I VGCAVCGEFKKA QVQEQLAR
                350        360        370        380        390        400

                        100         110         120         130         140         150
orf61.pep      KIEWLPSSAQAXGIRNH YRHPEEHGS DRWFN ALGSR RFSRNACVVVSCGTAVTVDALTDD
                ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 orf61a      KIEWLPSSAQA LGIRNH YRHPEEHGS DRWFN ALGSR RFSRNACVVVSCGTAVTVDALTDD
                410        420        430        440        450        460

                        160         170         180         189
orf61.pep      GHYLG XGTIM PGFHLMKES LAVRT ANLN RHAGKRYP FPT
                ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
55 orf61a      GHYLG-GTIMPG FHLMKES LAVRT ANLN RHAGKRYP FPTTT GNAVASGMMD AVCGSVM MM
                470        480        490        500        510        520

60 orf61a      HGRLKEKTGAGKPVDVIITGGGA AKVAEALPPAF LAENTVRVADNLVIHGL LNLIA EEG

```

530 540 550 560 570 580

The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```

1   ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
201 TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGTG TGACCCACCT
10 351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGGC GAAGTGGTCG CACCGTTTGG
401 GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTTTGACCG GCCGCGAGTAT
451 GAGTTGGGTT CGCTGTCGCC TGTTCGGCA GTGGCGTGCC GGCGCGCCTT
501 GTCGCGTTTG GGTTTGAAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
15 601 GCGAAAACGG TTGCGTGGT CCGTATCGGC ATCAATTTTCG TGCTGCCCAA
651 GGAAGTGGAA AACCGCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
701 GCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA
751 CTTGATGCGG TGTGTTGCA ATATGCGCGG GACGGATTTC CGCCTTTGT
801 GCGGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
20 851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
901 CAAGGCGTTC TGCATTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
951 CGCGGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCGGAGC
1001 GCGGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGGTGGAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
25 1101 GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGTGGATG
1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATT CAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGCGAGC GCCGCTTCAG CCGCAACGCC
30 1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCGCA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501 CGTTATCCTT TCCCAGCAC AACGGGCAAT GCCGTCGCCA CGGCGATGAT
1551 GGATGCGGTT TGCGGTCGGT TTATGATGAT GCACGGGCGT TTGAAAGAAA
35 1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACGGG CGGCGGCGCG
1651 GCAAAAGTTG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```

40 1   MTVLKPSHWR VLAELADGLP QHVSQARMAD MKPQQLNGF WQOMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLRELGER GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECIMFS FGWVFDPRQY
151 ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGRDKLG GILIETVRTG
45 201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLE
251 LDAVLLQYAR DGFAPFVAEY QANRDHGKA VLLLRDGETV FEGTVKGVDG
301 QGVLHLETAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
351 KWAWVENGTG ATVGSAFYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA
401 QVQEQLARKI EWLPSAQAL GIRNHYRHPE EHGS DRWENA LGSRRFSRNA
451 CUVVSCGTAV TVDALTDGHH YLGGTIMPGE HLMKESLAVR TANLNRHAGK
50 501 RYFPPTTTGN AVASGMMDAV CGSVMMHGR LKEKTGAGKP VDVIIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

```

ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```

55 orf61a.pep 10 20 30 40 50 60
    MTVLKPSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR
    |||||
orf61-1      10 20 30 40 50 60
    MTVLKLSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR

60 orf61a.pep 70 80 90 100 110 120
    LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
    |||||
orf61-1      70 80 90 100 110 120
    LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK

65 130 140 150 160 170 180

```

-174-

5	orf61a.pep	GRGRQGRKWSHRLGECLMFSFGWVFDPRQYELGSLSPVAAVACRRALSRLGLKTQIKWPN	
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVFDPRQYELGSLSPVAAVACRRALSRLGLDVQIKWPN	
		130 140 150 160 170 180	
10	orf61a.pep	DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	
	orf61-1	DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	
		190 200 210 220 230 240	
15	orf61a.pep	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGV	
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGV	
		250 260 270 280 290 300	
20	orf61a.pep	QGVLEHLETAEGKQTVVSGEISLRSDRPVSVPKRRDSEFLLLDGGNSRLKWAWVNGTF	
	orf61-1	QGVLEHLETAEGKQTVVSGEISLRSDRPVSVPKRRDSEFLLLDGGNSRLKWAWVNGTF	
		310 320 330 340 350 360	
25	orf61a.pep	ATVGSAPYRDLSPGLAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	
	orf61-1	ATVGSAPYRDLSPGLAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	
		370 380 390 400 410 420	
30	orf61a.pep	GIRNHYRHPEEHGSDRWFNALGSRRFNRNACVVVSCGTAVTVDALTDGHHYLGGTIMP	
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFNRNACVVVSCGTAVTVDALTDGHHYLGGTIMP	
		430 440 450 460 470 480	
35	orf61a.pep	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMHGRLEKKTGAGKP	
	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMHGRLEKKTGAGKP	
		490 500 510 520 530 540	
40	orf61a.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLI AEGGESEHTX	
	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMI AEGGREYEHIX	
		550 560 570 580 590	

Homology with a predicted ORF from *N.gonorrhoeae*ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*50 *gonorrhoeae*:

55	orf61.pep	EISLRSDXRPVSVXKRRDSEFLLLDGGNS	30
	orf61ng	TVCEGTVKGV DGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVKRPDSEFLLLEGGNS	211
60	orf61.pep	RLKWAWVNGTFATVGSAPYRDLSPGLAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKWAWVNGTFATVGSAPYRDLSPGLAEWAEKADGNVRIVGCAVCGESKKAQVQEQLAR	271
65	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFNRNACVVVSCGTAVTVDALTD	150
	orf61ng	KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFNRNACVVVSCGTAVTVDALTD	331
65	orf61.pep	GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT	189
	orf61ng	GHYLG-GTIMPGFHLMKESLAVRTANLNRPAKRYPFPTTTGNAVASGMMDAVCGSIMM	390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

```

      1 MFSFGWAFDR PQYELGSLSP VAALACRRAL GCLGLETQIK WPNDLVVGRD
      51 KLGGLIETV RAGGKTAVV GIGINFVLPK EVENAASVQS LFQTASRRGN
5    101 ADAAVLLET L LAELGAVLEQ YAEEGFAPFL NEYETANRDH GKAVLLLRDG
      151 ETVCEGTVKG VDGRGVHLHE TAEGEQTVVS GEISLRPDNR SVSVPKRPDS
      201 ERFLLEGGN SRLKWAVVEN GTFATVGSAP YRDLSP LGAE WAEKADGNVR
      251 IVGCAVCGES KKAQVKEQLA RKIEWLPSSA QALGIRNHYR HPEEHGSDRW
      301 FNALGSRRFS RNACVVVSCG TAVTVDALTD DGHYLGGTIM PGFHLMKESL
10   351 AVRTANLNRP AGKRYFPFTT TGNVAVSGMM DAVCGSIMMM HGRLKEKNGA
      401 GKPVDVIITG GGAAKVAEAL PPAFLAENTV RVADNLVIHG LLNLIAAEGG
      451 ESEHA*

```

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

```

      1 ATGACGGTTT TGAAGCCTTC GCATTGGCGG GTGTTGGCGG AGCTTGCCGA
15   51 CGGTTTGCCG CAACACGTAT CGCAATTGGC GCGTGAGGCG GACATGAAGC
      101 CGCAGCAGCT CAACGGTTT TGGCAGCAGA TGCCGCGCGA TATACGCGGG
      151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CTTGGCGGGT
      201 TTTGATGCC GAAAGTTTGC GCGATCTGGG GGAAAGGTCG GGTTTTCAGA
      251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
20   301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
      351 GCAAAGTAAG GGCAGGGGCG GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
      401 GCGAGTGCC TATGTTTCAGT TTCGGCTGGG CGTTTGACCG GCCGCGATAT
      451 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA CTTGCGTGCC GCGCGCTTTT
      501 GGGGTGTTTG GGTTTGGAAA CGCAAATCAA GTGGCCAAAC GATTTGCTCG
25   551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGGCGGGC
      601 GGTAAACCG TTGCCGTGGT CCGTATCGGC ATCAATTTCTG TGCTGCCCAA
      651 GGAAGTGGAA AACGCCGCTT CCGTGACGTC GCTGTTTCAG ACGGCATCGC
      701 GGCGGGGCAA TGCCGATGCC GCCGTATTGC TGGAAACATT GCTTGCGGAA
      751 CTGGGCGCGG TGTTGGAAACA ATATGCGGAA GAAGGGTTCG CGCCATTTTT
30   801 AAATGAGTAT GAAACGGCCA ACCGCGACCA CGGCAAGCGG GTATTGCTGT
      851 TGCGCGACGG CGAAACCGTG TCGAAGGCA CGGTTAAAGG CGTGGACGGA
      901 CGAGGCGTTC TGCACCTTGA AACGGCAGaa ggcgaACAGa cggcgtcag
      951 cggcgaaaTC AGcctGCggc ccgacaacaG GTCGGtttcc gtgcccgaagc
100  1001 ggcccgatTC GgaacgtTTT tTGctgttgg aaggcgggaa cagccgCTC
35   1051 AAGTGGGCGT GggtggAAaAa cggcacgttc gcaaccgtgg gcagcgcgCc
      1101 gtaCCGCGAT TTGTGCGCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
      1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATC CAAAAAGGCA
      1201 CAAGTGAAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCGCG
      1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
40   1301 CCGACCGTTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
      1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
      1401 TGACGGACAT TATCTCGGCG GAACCATCAT GCCCGGCTTC CACCTGATGA
      1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
      1501 CGTTACCCTT TCCCGACCAC AACGGGCAAC GCCGTCGCAA GCGGCATGAT
45   1551 GGACGCGGTT TGCGGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
      1601 AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CCGCGGCGCG
      1651 GCGAAAGTCG CCGAAGCCCT GCCGCCTGCA TTTTGGCGG AAAATACCGT
      1701 GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
      1751 CCGAAGGCGG GGAATCGGAA CACGCTTAA

```

50 This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

```

      1 MTVLKPSHRV VLAELADGLP QHVSQ LAREA DMKPOQLNGF WQOMP AHIRG
      51 LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQ TALKHEC ASSNDEILEL
55   101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWAFDRPQY
      151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILIETVRAG
      201 GKTAVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
      251 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVGD
      301 RGVHLLETAE GEQTVVSGEI SLRPNDRSVS VPKRPDSERF LLLEGGNSRL
      351 KWAVVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGESKKA
      401 QVKEQLARKI EWLPPSAQAL GIRNHYRHPE EHGS DRWFNA LGSRRF SRNA
60   451 CVVVS CGTAV TVDALTD DGH YLGGT IMPGF HLMKESLAVR TANLNR PAGK
      501 RYFPFTTGN AVASGMMDAV CGSIMMMHGR LKEKNGAGKP VDVIITGGGA
      551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA*

```

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

	orf61ng-1.pep	MTVLKPSHWRVLAELADGLPQHVSQALAREADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR	60
	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR	60
5	orf61ng-1.pep	LVRPLAVFDAEGLRDLGERSGFGTALKHECASSNDEILELARIAPDKAHTICVTHLQSK	120
	orf61-1	LVRPLAVFDAEGLRELGERSGFGTALKHECASSNDEILELARIAPDKAHTICVTHLQSK	120
10	orf61ng-1.pep	GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN	180
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAACRRALSRLGLDVQIKWPN	180
15	orf61ng-1.pep	DLVVGRDKLGGILITVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61-1	DLVVGRDKLGGILITVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
20	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG	300
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG	300
25	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPDNRVSVPKRPDSEFLLLEGGNSRLKQAWVENGTF	360
	orf61-1	QGVHLHLETAEGKQTVVSGEISLRSDDRVSVPKRRDSEFLLLDGGNSRLKQAWVENGTF	360
30	orf61ng-1.pep	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL	420
	orf61-1	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQQLARKIEWLPSSAQAL	420
35	orf61ng-1.pep	GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHLGGTIMGPF	480
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHLGGTIMGPF	480
40	orf61ng-1.pep	HLMKESLAVRTANLNRPAKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNAGKGP	540
	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMHGRLKEKTGAGKP	540
	orf61ng-1.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNIAAEGGESEHAX	593
	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMAAEGREYEHIX	593

Based on this analysis, including the homology with the baf protein of *B.pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 29

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 241>:

	1	ATGTTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
	101	GCCTGCTAAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCCC	CCGTATGTC
50	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCTG
	201	CAACTATGTG	CTGACCTGTC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTCGGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
	351	ATGCGGCGCG	GCGGCATTG	CCGTGTCGC	GCTGCTGATG	GCGGGCGGTG
55	401	CGGaAGAGGG	CGGCGaAGTC	GGCTGGTTG	GCTGCCTGCT	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTG	CGCTATGCGT	CCGACGCAAA	GGCTGATTG
	501	ACGCATCGGC	GCACCGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGCTG	TATTTGGGTT	TGGGGTGC..

60 This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

-177-

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
 51 GKIIPREEWKP LLIVSFVNYV LTLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 201 WSVGMVLSLL YLGLGC..

Further work revealed the complete nucleotide sequence <SEQ ID 243>:

1 ATGTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
 101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
 151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTTCGT
 201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
 251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
 301 TTTGTGCGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
 351 ATGCGGCGCG GCGGCATTTC CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
 401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
 451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
 551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTGCGGTT TGGGGTGCGG
 651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
 701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG
 751 GCGGTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG CCTTGGGCGT
 801 GTTTGTCGTC ATCGCCGCCA CTTGTTGTCG CGGCCGGCTG TCGCATCAAA
 851 AATAA

25 This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
 51 GKIIPREEWKP LLIVSFVNYV LTLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 201 WSVGMVLSLL YLGLGCGWYA YLWLNKMSR VPANVSGLLI SLEPVVGVLL
 251 AVLILGEHLS PVSAALGVFV IAATLVAGRL SHQK*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

35 Orf62 1 MFYQILALI WSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV GKIIPREEWKP 60
 M YQILAL+IWSSS I K Y +DP L+V VR R KI + K
 HI0976 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVQVRLIIAMIIVMPLFLRRWKKIDKPMRQ 60
 Orf62 61 LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAY 114
 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +
 40 HI0976 61 LWLAFNYTAVFLLQFVGLKYTSAASAVTMIGLEPLLVFVGHFFFKTKQNGF 114

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N.*

45 *meningitidis*:

10 20 30 40 50 60
 orf62.pep MFYQILALI WSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV GKIIPREEWKP
 |||||
 50 orf62a MFYQILALI WSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV GKIIPREEWKP
 |||||
 10 20 30 40 50 60
 orf62.pep LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA
 |||||
 55 orf62a LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA
 |||||
 70 80 90 100 110 120
 orf62.pep AAFAGVALLMAGGAEEGGEV GWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
 |||||
 130 140 150 160 170 180
 orf62a AAFAGVALLMAGGAEEGGEV GWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
 |||||

-178-

```

      |||
orf62a  AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
      130      140      150      160      170      180

5      190      200      210
orf62.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
      |||
orf62a  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVCWSYAYWLWNKGMSRVPANVSGLLI
      190      200      210      220      230      240

10      SLEPVVGVLAVLILGEHLSPVSVLGVFVVAATLVAGRLSHQKX
      250      260      270      280

```

The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

```

15      1  ATGTTTACC  AAATCCTTGC  CCTGATTATC  TGGAGCAGCT  CGTTTATTGC
      51  CGCCAAATAT  GTCTATGGCG  GCATCGATCC  CGCATTGATG  GTCGGCGTGC
      101  GCCTGCTGAT  TGCTGCGCTG  CCTGCACTGC  CCGCCTGCCG  CCGTCATGTC
      151  GGCAAGATTC  CGCGTGAGGA  ATGGAAGCCG  TTGCTGATTG  TGTCGTTGCT
      201  CAACTATGTG  CTGACCCTGC  TACTTCAGTT  TGTCGGGTTG  AAATACACTT
      251  CCGCCGCCAG  CGCATCGGTC  ATTGTTCGGAC  TCGAGCCACT  GCTGATGTG
20      301  TTTGTTCGGAC  ACTTTTCTT  CAACGACAAA  GCGCGTGCCT  ACCACTGGAT
      351  ATGCGGCGCG  GCGGCATTG  CCGGTGTGCG  GCTGCTGATG  GCGGGCGGTG
      401  CGGAAGAGGG  CGGCGAAGTC  GGCTGGTTCG  GCTGCCTGCT  GGTGTTGTTG
      451  CCGGGCGCGG  GCTTTTGTG  CGCTATGCGT  CCGACGCAAA  GGCTGATGTC
      501  ACGCATCGGC  GCACCGGCAT  TCACATCTGT  TTCCATTGCC  GCCGCATCGT
25      551  TGATGTGCT  GCCGTTTTCG  CTTGCTTTGG  CGCAAAGTTA  TACCGTGGAC
      601  TGGAGCGTCG  GAATGGTATT  GTCGCTGCTG  TATTGGGCG  TGGGGTGCAG
      651  CTGGTACGCC  TATTGGCTGT  GGAACAAGGG  GATGAGCCGT  GTTCCTGCCA
      701  ACGTTTCGGG  ACTGTTGATT  TCGCTCGAAC  CCGTCGTCGG  CGTGCTGCTG
      751  GCGGTTTGA  TTTTGGGCGA  ACACCTGTCG  CCCGTGTCCG  TCTTGGGCGT
30      801  GTTTGTGCTC  ATCGCCGCCA  CTTGTTGTC  CGGCCGGCTG  TCGCATCAAA
      851  AATAA

```

This encodes a protein having amino acid sequence <SEQ ID 246>:

```

35      1  MFYQILALII  WSSSFIAAKY  VYGGIDPALM  VGVRLIIAAL  PALPACRRHV
      51  GKIPREEWKP  LLIVSFVNYV  LTLQLQFVGL  KYTSAASASV  IVGLEPLLMV
      101  FVGHFFNDK  ARAYHWICGA  AAFAGVALLM  AGGAEEGGEV  GWFGCLLVLL
      151  AGAGFCAAMR  PTQRLIARIG  APAFTSVSIA  AASLMCLPFS  LALAQSYTVD
      201  WSVGMVLSLL  YLGVCWSYA  YWLWNKGMSR  VPANVSGLLI  SLEPVVGVL
      251  AVLILGEHLS  PVSVLGVFVV  IAATLVAGRL  SHQK*

```

ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:

```

40      orf62a.pep  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV 60
      orf62-1      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV 60

      orf62a.pep  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA 120
45      orf62-1      LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA 120

      orf62a.pep  AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180
50      orf62-1      AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180

      orf62a.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVCWSYAYWLWNKGMSRVPANVSGLLI 240
      orf62-1      AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI 240

55      orf62a.pep  SLEPVVGVLAVLILGEHLSPVSVLGVFVVAATLVAGRLSHQKX 285
      orf62-1      SLEPVVGVLAVLILGEHLSPVSVLGVFVVAATLVAGRLSHQKX 285
60

```

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N. gonorrhoeae*:

	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGIKIPREEWKP	60
	orf62ng	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGIKIPREEWKP	60
5	orf62.pep	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA	120
10	orf62.pep	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA	180
	orf62ng	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA	180
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC	216
15	orf62ng	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANASGLLI	240

The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

	1	ATGTTTACC	AAATCCTGC	CCTGATTATC	TGGGGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
20	101	GCCTGCTGAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCGG	CCGTCTATGC
	151	GGCAAGATTG	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCTG
	201	CAACTATGTG	CTGACCTGTC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTGCGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCC	ACCCTGGAT
	351	ATGCGGCGCG	GCGGCATTG	CCGGTGTCGC	GCTGCTGATG	GCGGGCGGTG
25	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTTCG	GCTGCCTGCT	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	CCGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGTTG	TATTGGGTT	TGGGGTGGCG
30	651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTTCTGCCA
	701	ACGCGTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCGG	CGTGCTGTTG
	751	GCGGTTTGA	TTTTGGGCGA	ACATTTATCG	CCCGTGTCG	CCTTGGGCGT
	801	GTTTGTGCTC	ATCGCCGCCA	CTTTCGCCGC	CGGCCGGCTG	TCGCGCAGGG
	851	ACGCGCAAAA	CGGCAATGCC	GTCTGA		

35 This encodes a protein having amino acid sequence <SEQ ID 248>:

	1	MFYQILALII	WGSSSFIAKY	VYGGIDPALM	VGVRLLIAAL	PALPACRRHV
	51	GKIPREEWKP	LLIVSFVNYV	LTLLQLFVGL	KYTSAAASV	IVGLEPLLMV
	101	FVGHFFNDK	ARAYHWICGA	AAFAGVALLM	AGGAEEGGEV	GWFGCLLVLL
40	151	AGAGFCAAMR	PTQRLIARIG	APFTSVSIA	AASLMCLPFS	LALAQSYTVD
	201	WSVGMVLSLL	YLGLGCGWYA	YWLWNKGMSR	VPANASGLLI	SLEPVGVLL
	251	AVLILGEHLS	PVSALGVFVV	IAATFAAGRL	SRRDAQNGNA	V*

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

		10	20	30	40	50	60
45	orf62ng.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGIKIPREEWKP					
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGIKIPREEWKP					
		10	20	30	40	50	60
50	orf62ng.pep	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA					
	orf62-1	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA					
		70	80	90	100	110	120
55	orf62ng.pep	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA					
	orf62-1	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA					
		130	140	150	160	170	180
60	orf62ng.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANASGLLI					
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANVSGLLI					
65		190	200	210	220	230	240

-180-

```

                250      260      270      280      290
orf62ng.pep    SLEPVVGVLAVLILGEHLSPVSALGVFVVIATFAAGRLSRRDAQNGNAVX
5             orf62-1    SLEPVVGVLAVLILGEHLSPVSALGVFVVIATLVAGRLSHQKX
                250      260      270      280

```

Furthermore, ORF62ng shows significant homology to a hypothetical *H. influenzae* protein:

```

10  sp|Q57147|Y976_HAEIN HYPOTHETICAL PROTEIN HI0976 >gi|1074589|pir||B64163
    hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
    >gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
    Score = 106 bits (262), Expect = 2e-22
    Identities = 56/114 (49%), Positives = 68/114 (59%)

15  Query: 1  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXXXCRRHVKGKIPREEWKP 60
    M YQILAL+IW SS I K Y +DP L+V VR R KI + K
    Sbjct: 1  MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60

    Query: 61  LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAY 114
    L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +
20  Sbjct: 61  LWLWLAFFNYTAVFLLQLFGLKYTSAASAVTMIGLEPLLVVVGHFFFKTKQNGF 114

```

Based on this analysis, including the homology with the transmembrane protein of *H. influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 30

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 249>:

```

30  1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkgTA
    51  sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTCTT
    101 GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
    151 TTGGCAGGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
    201 CGGTTTCGtA srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
    251 GCCGkACTGC CCGGCGTGTT TCTGTTCCGGC TTTCCCGCAC AGTTCATCAA
    301 CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTTGAAC
35  351 GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
    401 CTCGGCAACG CCGTCCCCGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
    451 GCCCGGGGAT ATGGGCAGGG TGCTGGAACA TTACGCCGCG AGCGGTTTTC
    501 CCCAGCTTGC CCTGTACAAy ksCGCAAGCG GCAAAATCGA AAAAAGCATC
    551 AACCCGCACA AGCTCGATCA GCCGTTTCCA GGTAAGGCGC GTTGGGAaAa
40  601 AATCCaACGG GCGGGTTTCG TCAGGGATTG GGAAGGCATA GCGGCGGTAT
    651 TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACwACGG GCGCGATTAC
    701 GCCTTGTTTT TCCGTCAGCC GGTTCCTCAA GGCCTGGCAG AGGATGCCGT
    751 yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
45  801 AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
    851 TCGATTTTTC TTGCACTGGT CATGGCACTG TATTTTCGCC GCGGTTTCGT
    901 CGAACCCGTC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
    951 ATTTACGCCA GACGCGCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
    1001 ACCArGTTGT TCAACCACAT GACCGAGCAG CTTTCCATCG CCAAAGATGC
    1051 AGACGAGCGC AACCGCCGGC GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
50  1101 GCGTGTGGA GGGGCTGACC ACGGGCGTGG TGGTGTTTGA CGAACAAGGC
    1151 TGTCTGAAAA CCTTCAACAA AGCGGCGGGT ACC..

```

This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

```

55  1  MRRFLPIAAI CAXXLXXGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
    51  LARYVILLK DRRDGVFGSX XAKXPXXXMF TLVAXLPGVF LFGFPAQFIN
    101 GTINSWFGND THEALERSLN LSKSALNLAA DNALGNAVVP QIDLIGAASL
    151 PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PFPKGARWEK
    201 IORAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFFRQP VPKGVAEDAV
    251 LIEKARAKYA ELSYSKKGLO TFFLATLLIA SLLSIFLALV MALYFARRFV

```

-181-

301 EPVLSLAEGA KAVAQGDfsQ TRPVLrNDEF GRLTXLFNHM TEQLSIAKDA
 351 DERNRREEA ARHYLECVLE GLTTGVVVFDE EQGCLKTFNK AAGT..

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

```

      1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
5      51 CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
     101 GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTG GTCCGCCGTT
     151 TTGGCACGTT ATGTCATATT GCTGTGAAA GACAGGCGCG ACGGCGTATT
     201 CGGTTCGCAG ATTGCCAAAC GCCTTTCTGG GATGTTTACG CTGGTTGCCG
     251 TACTGCCCGG CGTGTTTCTG TTCGGCGTTT CCGCACAGTT CATCAACGGC
10     301 ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TGAACGCAG
     351 CCTCAATTG AGCAAGTCCG CATTGAATTT GCGGCAGAC AACGCCCTCG
     401 GCAACGCCGT CCCCCTGCAG ATAGACCTCA TCGGCGCGGC TTCCTGCCCC
     451 GGGGATATGG GCAGGGTGCT GGAACATTAC GCGGCAGCG GTTTTGCCCA
     501 GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCAACGACC
15     551 CGCACAAAGCT CGATCAGCCG TTTCAGGTA AGGCGCGTTG GGAAAAATC
     601 CAACGGGCGG GTTCGGTCAG GGATTGGAA AGCATAGGCG GCGTATTGTA
     651 CGCGCAGGGC TGGCTGTCGG CGGGTACGCA CAACGGGCGC GATTACGCCT
     701 TGTTTTTCCG TCAGCCGTTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
     751 ATCGAAAAGG CAAGGGCGAA ATATGCTGAG TTGAGTTACA GCAAAAAGG
20     801 TTTGCAGACC TTTTTCCTGG CAACCTGCT GATTGCCTCG CTGCTGTCGA
     851 TTTTCTTGC ACTGGTCATG GCACTGTATT TCGCCCGCGG TTTCTGCGAA
     901 CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATTT
     951 CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
25    1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
    1051 GAGCGCAACC GCGGCGCGGA GGAAGCCGCC AGGCATTATC TTGAATGCGT
    1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
    1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
    1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
    1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACCG
30    1301 ACAAAACCGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
    1351 CTGGGCAAGG CAACCGTCTT GCCCGAAGAC AACGGCAACG GCGTGTAAT
    1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
    1451 GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCAGC
    1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
35    1551 GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
    1601 AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCCG CAATTATGCG
    1651 CGTTCCCCTT CGCTCAAATT GGAAATCAG GATTTGAACG CCTTAATCGG
    1701 CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
    1751 TTGCCGGCGA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCCGAG
40    1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
    1851 TGTGCCCGAA GTCAGGTTAA AATCGGAAAC AGGGCAGGAC GGTCCGATTG
    1901 TCCTGACGGT TTGCGACAAC GGCAAAGGGT TCGGCAGGGA AATGCTGCAC
    1951 AACGCCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG
    2001 TCTGCCTGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
45    2051 TGAGCAATCA GGATGCGGTT GCGCGTGTG TCAGAATCAT CTTGCCAAAA
    2101 ACGGTAAAAA CTTATGCGTA G
  
```

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

```

      1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
50     51 LARYVILLLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
     101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAPVQ IDLIGAASLP
     151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
     201 QRAGSVRDLE SIGGVLYAQQ WLSAGTHNGR DYALFFRQPV PKGVAEDAVL
     251 IEKARAKYAE LSYSKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFE
55     301 PVLSLAEGAK AVAQGDfsQT RPVLrNDEF GRLTXLFNHMT EQLSIAKEAD
     351 ERNRREEA ARHYLECVLE LTGTVVVFDE QGCLKTFNK AEQILGMPLT
     401 PLWSSRHGW HGVSAQQSLL AEFVFAIGAA AGTDKPVHVK YAAPDDAKIL
     451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
     501 PIQLSAERLA WKLGGKLDEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA
     551 RSPSLKENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLT VAADTTAMRO
60     601 VLHNIKFNAE EAAEEADVPE VRVKSETGQD GRIVLTVCDN KGKFGREMLH
     651 NAFEPYVTDK PAGTGLGLPV VKKIIIEHGG RISLSNQDAG GACVRIILPK
     701 TVKTYA*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N. meningitidis*:

5	orf64.pep	10 20 30 40 50 60	MRRFLPIAAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLLK
	orf64a	10 20 30 40 50 60	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLLK
10	orf64.pep	70 80 90 100 110 120	DRRDGVFGSXXAKXPXXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN
	orf64a	70 80 90 100 110 120	DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN
15	orf64.pep	130 140 150 160 170 180	LSKSALNLAADNALGNAPVQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNXASGKIE
	orf64a	120 130 140 150 160 170	LSKSALNLAADNALGNAIPVQIDXIGAASLPDGMGRVLEHYAGSGFAQLALYNAASGKIE
20	orf64.pep	190 200 210 220 230 240	KSINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP
	orf64a	180 190 200 210 220 230	KSINPHKLDQFPFGKARWEKIQAGSVRDLESIGGVLYAXGWLSAXTHNGRDYALFFRQP
30	orf64.pep	250 260 270 280 290 300	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
	orf64a	240 250 260 270 280 290	VPKGVAEDAVLIEKARAXXXLSYSKKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
35	orf64.pep	310 320 330 340 350 360	EPVLSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTXLFNHMTEQLSIAKDADERNRRREEA
	orf64a	300 310 320 330 340 350	EPVLSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEA
40	orf64.pep	370 380 390	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT
	orf64a	360 370 380 390 400 410	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSL
45	orf64a	420 430 440 450 460 470	LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQ

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

50	1	ATGCGCCGTT	TTCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGTTGTA
	51	CGGACTGACG	GCGGCAACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTTCT
	101	GGTGGATTGT	TGCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCG	ACGGCGTATT
	201	CGGTTTCGAG	ATTGCCAAC	GCCTTTCGGG	GATGTTTACG	CTGGTTGCCG
55	251	TACTGCCCGG	CTGTGTTCTG	TTCGGCGTTT	CCGCACAGTT	TATCAACGGC
	301	ACGATTAATT	CGTGGTTCGG	CAACGATACC	CACGAGGCGC	TTGAACGCAG
	351	CCTCAATTG	AGCAAGTCCG	CATTGAATCT	GGCGGCAGAC	AACGCCCTTG
	401	CAACGCCAT	CCCCGTGCAG	ATAGACNTCA	TCCGCGCGGC	TTCCCTGCC
	451	NGGGATATGG	GCAGGGTGCT	GGAACATTAC	GCCGGCAGCG	GTTTTGCCCA
60	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGCAA	AATCGAAAAA	AGCATCAACC
	551	CGCACAAGCT	CGATCAGCCG	TTTCCAGGTA	AGGCGCGTTG	GGAAAAATC
	601	CAACAGGCGG	GTTCCGTGAG	GGATNNGGAA	AGCATAGGCG	GCGTATTGTA
	651	CGCGCANGGC	TGGCTGTCGG	CAGNNACGCA	CAACGGGCGC	GATTACGCCT
	701	TGTTTTTCCG	TCAGCCGGTT	CCCAAAGGCG	TGGCAGAGGA	TGCCGTCTTA
65	751	ATCGAAAAGG	CAAGGGCGNA	ANANNNTNAG	TTGAGTTACA	GCAAAAAGG
	801	TTTGACAGAC	TTTTTCCTNG	CAACCCTGCT	GATTGCCTCN	CTGCTGCGA
	851	TTTTTCTTGC	ACTGGTCATG	GCACTGTATT	TCGCCGCCG	TTTCGTGAA

-183-

5
 10
 15
 20
 25

```

901 CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATT
951 CAGCCAGACG CGCCCCGTGT TGCACAACGA CGAGTTCGGA CGCTTGACCA
1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGACATTATC TCGAATGCGT
1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
1151 TGAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 TCCCTTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301 ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAATCCTG
1351 CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACNGCAACG GCGTGGAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCTCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAGGCT
1551 GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
1601 AACAAAGTGGC GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG
1651 CGTTCCTCTT CGNCTCAATT GGAAAATCAG GATTGAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGGTTT GCGGCGGAAC
1751 TTGCCGGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TGTGCCCCGAA GTCAGGGTAA AATCGGAAGC GGGGCAGGAC GGACGGATTG
1901 TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCAGGGA AATGCTGCAC
1951 AATGCCTTCG AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG
2001 ACTGCCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC
2051 TGAGCAATCA GGATGCGGGC GCGCGTNTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G
  
```

This encodes a protein having amino acid sequence <SEQ ID 254>:

30
 35
 40

```

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QQAGSVRDXE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301 PVLSLAEGAK AVAQGDFSQT RPVLNRNDEF RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401 PLWGSSRHGW HGVSAQQSLL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL
451 LGKATVLPED NXNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDX DAQILTRSTD TIIKQVAALK EMVEAFRNYX
551 RSPSXQLENO DLNALIGDVL ALYEAGPCR FAEELAGEPLM MAADTTAMRQ
601 VLHNIFKNAA EAAEEADVPE VRVKSEAGQD GRIVLTVC DN KKGFGREMLH
651 NAFEPYVTDK PAGTGLXLPV VKKIIIEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*
  
```

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

45
 50
 55
 60
 65

```

              10      20      30      40      50      60
orf64a.pep  MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAML
|||||
orf64-1      MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAML
              10      20      30      40      50      60

              70      80      90      100     110     120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
|||||
orf64-1      DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              70      80      90      100     110     120

              130     140     150     160     170     180
orf64a.pep  SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
|||||
orf64-1      SKSALNLAADNALGNAIPVQIDLIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              130     140     150     160     170     180

              190     200     210     220     230     240
orf64a.pep  SINPHKLDQFPFGKARWEKIQQAGSVRDXE SIGGVLYAXGWLSAXTHNGR DYALFFRQPV
|||||
orf64-1      SINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGR DYALFFRQPV
              190     200     210     220     230     240

              250     260     270     280     290     300
  
```

-184-

5	orf64a.pep	PKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
10	orf64a.pep	PVLSLAEGAKAVAQGDQFSQTRPVLNRNDEFGRLTKLFNHNMTQLSIAKEADERNRRREEAA
	orf64-1	PVLSLAEGAKAVAQGDQFSQTRPVLNRNDEFGRLTKLFNHNMTQLSIAKEADERNRRREEAA
15	orf64a.pep	RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTFLWGSSRHGWGVSQAQSSLL
	orf64-1	RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTFLWGSSRHGWGVSQAQSSLL
20	orf64a.pep	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNNGVVMVIDDITVLIHAQK
	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNNGVVMVIDDITVLIHAQK
25	orf64a.pep	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDITIKQVAALK
	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDITIVKQVAALK
30	orf64a.pep	EMVEAFRNYXRSPSXQLENQDLNALIGDVLALYEAGPCRFAELAGEPLMMAADTTAMRQ
	orf64-1	EMVEAFRNYARSPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLTVAADTTAMRQ
35	orf64a.pep	VLHNI FKNAAEAAEEADVPEVRVKSEAGQDGRIVLTVCDNGKGFGREMLHNAFEPVYTDK
	orf64-1	VLHNI FKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGREMLHNAFEPVYTDK
40	orf64a.pep	PAGTGLXLPVVKKIIIEHGGXISLSNQDAGGAXVRIILPKTVETYAX
	orf64-1	PAGTGLGLPVVKKIIIEHGGRIISLSNQDAGGACVRIILPKTVKTYAX

Homology with a predicted ORF from *N.gonorrhoeae*ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N.*50 *gonorrhoeae*:

55	orf64.pep	MRRFLPIAAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAML LVL SAVLARYVILL LK	60
	orf64ng	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAML LVL SAVLARYVILL LK	60
60	orf64.pep	DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	120
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
65	orf64.pep	LSKSALNLAADNALGNAPVPQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE	180
	orf64ng	LSKSALDLAADNAVSNAPVPQIDLIGTASLSGNMGVLEHYAGSGFAQLALYNAASGKIE	179
65	orf64.pep	KSINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
	orf64ng	KSINPHQFDQPLPDKEHWEIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239

-185-

	orf64.pep	VPKGVAEDAVLIEKARAKYAEISYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	300
	orf64ng	IPENVAQDAVLIEKARAKYAEISYSKKGLQTFFLVTLIASLLSIFLALVMALYFARRFV	299
5	orf64.pep	EPVLSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKDADERNRRREEA	360
	orf64ng	EPILSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTKLFNHMTQLSIAKEADERNRRREEA	359
10	orf64.pep	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT	394
	orf64ng	ARHYLECVLDGLTTGVVVSYP LSCCRTAVFSTCHSSPLSYF	400

An ORF64ng nucleotide sequence <SEQ ID 255> was predicted to encode a protein having amino acid sequence <SEQ ID 256>:

	1	MRRFLPIAAI	CAVVLLYGLT	AATGSTSSLA	DYFWWIVSFS	AMLLLVL SAV
15	51	LARYVILLK	DRRNGVFGSQ	IAKRLSGMFT	LVAVLPGLFL	FGISAQFING
	101	TINSWFGNDT	HEALERSLNL	SKSALDLAAD	NAVSNAVPVQ	IDLIGTASLS
	151	GNMGSVLEHY	AGSGFAQLAL	YNAASGKIEK	SINPHQFDQP	LPDKEHWEQI
	201	QQTGSVRSLE	SIGGVLYAQG	WLSAGTHNGR	DYALFFRQPI	PENVAQDAVL
	251	IEKARAKYAE	LSYSKKGLQT	FFLVTLIAS	LLSIFLALVM	ALYFARRFVE
20	301	PILSLAEGAK	AVAQGDIFSQT	RPVLNRNDEF	RLTKLFNHMT	EQLSIAKEAD
	351	ERNRRREEAA	RHYLECVLDG	LTTGVVVSYP	LSCCRTAVFS	TCHSSPLSYF*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

	1	ATGCGCCGCT	TCCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGCTGTA
	51	CGGATTGACG	GCGGCGACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTTCT
25	101	GGTGGATAGT	CTCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCA	ACGGCGTGTT
	201	CGGTTTCGCG	ATTGCCAAAC	GCCTTTCCGG	GATGTTACAG	CTGGTCGCCG
	251	TACTGCCCGG	CTTGTTCTCG	TTCCGCAATT	CCGCGCAGTT	TATCAACGGC
	301	ACGATTAATT	CGTGGTTCGG	CAACGACACC	CACGAAGCCC	TGAACGCAG
30	351	CCTTAATTTG	AGCAAAGTCCG	CACTGGATTT	GGCGGCAGAC	AATGCCGTCA
	401	GCAACGCCGT	TCCCGTACAG	ATAGACCTCA	TCGGCACCGC	CTCCCTGTCT
	451	GGCAATATGG	GCAATGTGCT	GGAACACTAC	GCCGGCAGCG	GTTTGGCCCA
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGGAA	AATCGAAAAA	AGCATCAATC
	551	CGCACCAATT	CGACCAGCCG	CTTCCCGACA	AAGAACATTG	GGAACAGATT
35	601	CAGCAGACCG	GTTCGGTTCG	GAGTTTGGA	AGCATAGGCG	GCGTATTGTA
	651	CGCGCAGGGA	TGGTTGTCCG	CAGGTACGCA	CAACGGGCGC	GATTACGCGC
	701	TGTTCTTCCG	CCAGCCGATT	CCCGAAAATG	TGGCACAGGA	TGCCGTCTCT
	751	ATTGAAAAGG	CGCGGGCGGA	ATATGCCGAA	TTGAGTTACA	GCAAAAAAGG
	801	TTTGACAGAC	TTTTTCTGCT	TAACCTGTCT	GATTGCTCTG	CTGCTGTCTG
40	851	TTTTTCTTGC	GCTGTAAATG	GCACTGTATT	TTGCCGCGCG	TTTCGTCTGAA
	901	CCCATTCTGT	CGCTTGCCGA	GGGCGCAAAG	GCGGTGGCGC	AGGTGTATTT
	951	CAGCCAGACG	CGCCCGGTAT	TGCGCAACGA	CGAGTTCGGA	CGTTTGACCA
	1001	AGCTGTTCAA	CCATATGACC	GAGCAGCTTT	CCATCGCCAA	AGAAGCAGAC
	1051	GAACGCAACC	GCCGGCGCGA	GGAAGCCGCC	CGTCACTACC	TCGAGTGCGT
45	1101	GTTGGATGGG	TTGACTACCG	GTGTGGTGGT	GTTTGACGAA	AAAGGCCGTT
	1151	TGAAAACCTT	CAACAAGGCG	GCGGAACAGA	TTTTGGGGAT	GCCGCTCGCC
	1201	CCCCTGTGGG	GCAGCAGCCG	GCACGGTTGG	CACGGCGTTT	CGGCGCAGCA
	1251	GCCCTGTGCT	GCCGAAGTGT	TtgccgccAT	CGGTGCGGCG	GCAGGTACGG
	1301	ACAAACCGGT	CCAGGTGGAA	TATGCCGCGC	CGGACGATGC	CAAAATCCTG
50	1351	CTGGGCAAGG	CGACGGTATT	GCCCGAAGAC	AACGGCAACG	GCGTGGTGAT
	1401	GGTGATTGAC	GACATCACCG	TGCTGATACG	CGCGCAAAAA	GAAGCCGCGT
	1451	GGGGTGAAGT	GGCGAAGCGG	CTGGCACACG	AAATCCGCAA	TCCGCTCACG
	1501	CCCATCCAGC	TTTCCGCCGA	ACGGCTGGCG	TGGAATTTGG	GCGGGAAGCT
	1551	GGACGATCAG	GACGCGCAAA	TCCTGACGCG	TtcgACCGAC	ACCATCATCA
55	1601	AACAGgtggc	gGCGTTAAAA	GAAATGGTCG	AGGCATTCCG	CAATTACGCG
	1651	CGCGCCCTT	CGCTCAAAC	GGAAAATCAG	GATTTGAACG	CCTTAATCGG
	1701	CGATGTTTTG	GCCCTGTACG	AAGCCGGCCC	GTGCCGGTTT	GAGGCGGAAC
	1751	TTGCCGGCGA	ACCGCTGATG	ATGGCGGCGG	ATACGACCGC	CATGCGGCAG
	1801	GTGCTGCACA	ATATTTTCAA	AAATGCCGCC	GAAGCGGCGG	AAGAAGCCGA
60	1851	TATGCCCGAA	GTCAGGTTAA	AATCGGAAAC	GGGGCAGGAC	GGACGGATTG
	1901	TCCTGACGGT	TTGCGACAAC	GGCAAGGGAT	TCGGCAAGGA	AATGCTGCAC
	1951	AATGCTTTTC	AGCCGTATGT	GACGGATAAG	CCGGCGGGAA	CGGGACTGGG
	2001	TCTGCCTGTA	GTGAAAAAAA	TCATTGGAGA	ACACGGCGGC	CGCATAGGCC
	2051	TGAGCAATCA	GGATGCGGGT	GGGGCGTGTG	TCAGAAATCAT	CTTGCCAAAA
65	2101	ACGGTAGAAA	CTTATGCGTA	G		

This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

```

      1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV
    51 LARYVILLLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING
   101 TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVVPQ IDLIGTASLS
    5 151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQP LPDKEHWEQI
   201 QQTGSVRSLE SIGGVLYAQQ WLSAGTHNGR DYALFFRQPI PENVAQDAVL
   251 IEKARAKYAE LSYSKKGLQT FFLVTLIIAS LLSIFLALVM ALYFARRFVE
   301 PILSLAEGAK AVAQGDFSQT RPLVRNDEFG RLTKLFNHMT EQLSIAKEAD
   351 ERNRRREEAA RHYLECVLDG LTGTVVVFDE KGRCLKTFNKA AEQILGMPLA
   10 401 PLWGSSRHGW HGVSAQQSLL AEFVFAAIGAA AGTDKPVQVE YAAPDDAKIL
   451 LGKATVLPED NGNGVVMVID DITVLIRAQK EAAWGEVAKR LAHEIRNPLT
   501 PIQLSAERLA WKLGGKLDQ DAQILTRSTD TTIKQVAALK EMVEAFRNYA
   551 RAPSLKLENQ DLNALIGDVL ALYEAGPCRF EAELAGEPLM MAADTTAMRQ
   601 VLNHIFKNAA EAAEEADMPV VRVKSETGQD GRIVLTVCDN KGKFGKEMLH
   15 651 NAFEPYVTDK PAGTGLGLPV VKKIIGEHGG RISLSNQDAG GACVRIILPK
   701 TVETYA*

```

ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

```

      10      20      30      40      50      60
  orf64ng-1.pep MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVSFSAMLLLVLSAVLARYVILLK
  20 orf64-1      MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAMLLLVLSAVLARYVILLK
      10      20      30      40      50      60

      70      80      90      100     110     120
  orf64ng-1.pep DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
  25 orf64-1      DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLGVSAQFINGTINSWFGNDTHEALERSLNL
      70      80      90      100     110     120

      130     140     150     160     170     180
  orf64ng-1.pep SKSALDLAADNAVSNVAVPQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
  30 orf64-1      SKSALNLAADNALGNVAVPQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK
      130     140     150     160     170     180

      190     200     210     220     230     240
  orf64ng-1.pep SINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQQWLSAGTHNGRDYALFFRQPI
  35 orf64-1      SINPHKLDQPPGKARWEKIQRAGSVRDLESIGGVLYAQQWLSAGTHNGRDYALFFRQPV
      190     200     210     220     230     240

      250     260     270     280     290     300
  orf64ng-1.pep PENVAQDAVLIEKARAKYAE LSYSKKGLQTTFFLVTLIIASLLSIFLALVMALYFARRFVE
  45 orf64-1      PKGVAEDAVLIEKARAKYAE LSYSKKGLQTTFFLATLLIASLLSIFLALVMALYFARRFVE
      250     260     270     280     290     300

      310     320     330     340     350     360
  orf64ng-1.pep PILSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA
  50 orf64-1      PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA
      310     320     330     340     350     360

      370     380     390     400     410     420
  orf64ng-1.pep RHYLECVLDGLTTGTVVVFDEKGRCLKTFNKA AEQILGMPLAPLWGSSRHGW HGVSAQQSLL
  55 orf64-1      RHYLECVLEGLTTGTVVVFDEQGCCLKTFNKA AEQILGMPLTPLWGSSRHGW HGVSAQQSLL
      370     380     390     400     410     420

      430     440     450     460     470     480
  orf64ng-1.pep AEFVFAAIGAAAGTDKPVQVEYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIRAQK
  60 orf64-1      AEFVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK
      430     440     450     460     470     480

      490     500     510     520     530     540
  orf64ng-1.pep EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTD TTIKQVAALK
  65 orf64-1      EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTD TTIKQVAALK

```


5	orf64-1	EAAGGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDEQDAQILTRSTDITIVKQVAALK	490	500	510	520	530	540
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGEPLMMAADTTAMRQ	550	560	570	580	590	600
10	orf64-1	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAAEELAGEPLTVAADTTAMRQ	550	560	570	580	590	600
	orf64ng-1.pep	VLHNIFKNAAEAAEEADMPVVRVKSETGQDGRIVLTVCDNGKGFGEMLHNAFEPYVTDK	610	620	630	640	650	660
15	orf64-1	VLHNIFKNAAEAAEEADMPVVRVKSETGQDGRIVLTVCDNGKGFGEMLHNAFEPYVTDK	610	620	630	640	650	660
	orf64ng-1.pep	PAGTGLGLPVVKKIIEHGGRISLSNQDAGGACVRIILPKTVETYAX	670	680	690	700		
20	orf64-1	PAGTGLGLPVVKKIIEHGGRISLSNQDAGGACVRIILPKTVKTYAX	670	680	690	700		

Furthermore, ORF64ng-1 shows significant homology to a protein from *A. caulinodans*:

25	sp Q04850 NTRY_AZOCA NITROGEN REGULATION PROTEIN NTRY >gi 77479 pir S18624 ntry protein - Azorhizobium caulinodans >gi 38737 (X63841) NtrY gene product [Azorhizobium caulinodans] Length = 771 Score = 218 bits (550), Expect = 7e-56 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)							
	Query: 7	IAAICAVVLLYGLTAATGSTSSLADYFWWIXXXXXXXXXXXXXXXXXXRYVILLKDRRNGV	66					
30	Sbjct: 35	ISALATFLILMGLTPVVPVTHQVVIS----VLLVNAAVLILSAMVGREIWRIRAKARARGR	90					
	Query: 67	FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKSALD	126					
35	Sbjct: 91	AAARLHIRIVGLFAVVSVPAILVAVVASLTLDRLDRWFMSMRTQEIIVASSVSAQTYVR	150					
	Query: 127	LAADNAVSNAVVPQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP	184					
40	Sbjct: 151	EHALNIRGDILAMSADLTRLKSV-----YEGDRSRFNQILTAQAALRNLPGLAMLI	200					
	Query: 185	HQFDQPLPDKEHWEIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA-----	233					
45	Sbjct: 201	RR-DLSVVERAN-VNIGREFIVPANLADGATPDQPVIIYLP--NDADYVAAVVPLKDYDD	256					
	Query: 234	--LFFRQPIPENVAQDAVLIEKARAKYAELSYSKGLQTFVLVXXXXXXXXXXXXXVMA	291					
50	Sbjct: 257	LYLYVARLIDPRVIGYKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWL	316					
	Query: 292	LYFARRFVEPILSLAEGAKAVAQGDQFSQTRPVLRLND-EFGRLTKLFNMHTEQLSIXXXXX	350					
55	Sbjct: 317	LNFSKWLVAPIRIRLMSAADHVAEGNLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI	376					
	Query: 351	XXXXXXXXXXXXHYLECVDGLTTGVVVFDEKGRKLTFNKAAEQILGMPLAPLWSSRHGW	410					
60	Sbjct: 377	LTARDQIDSRRRFTEAVLSGVGAGVIGLDSQERITILNRSERLLG--LSEVEALHRHLA	434					
	Query: 411	HGVSAQQSLLAEVFXXXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM	467					
65	Sbjct: 435	EVVPETAGLLEEA-----EHARQSVQGNITLTRDGRERVFAVRVTTEQSPEAHEGWVV	488					
	Query: 468	VIDDITVLIRAQKEAAAGGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDDQDAQILTR	527					
70	Sbjct: 489	TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFRGHV-TQDREIFDQ	547					
	Query: 528	STDITIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGE	587					
75	Sbjct: 548	CTDTIIRQVGDIGRMVDEFSSFARMKPKVVDSDQMSEIIRQTVFLMRVGHPEVVFDESEVP	607					
	Query: 588	PLMMAA-DTTAMRQVLHNI FKNXXXXXXXXXDMPEVRVK-----SETGQDGRIVLTVCD	639					
80	Sbjct: 608	PAMPARFDRRLVSQALTNILKNAAEAEIAPV-PDVRGQGRIRVSANRVGED--LVIDIID	664					

Query: 640 NGKGFGEKMLHNAFEPYVTDKPGATGLGLPVVKIIGEHHGRISLSNQDAG-GACVRIIL 698
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L
 Sbjct: 665 NGTGLPQESRNLLEPYVTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

```

1   ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCCGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCCGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG
451 CACGCGTTGG ATACG...
```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```

1   MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTPWAFSEF
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLRKAWW IAPNASTVIG
151 HALDT...
```

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

```

1   ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCCGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCCGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC
451 AAGCCTTGG ATACGCTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
501 CGATGGATT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCCG CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC
651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA
```

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

```

1   MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTPWAFSEF
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLRKAWW IAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGI AFVDYLFKLTVC TLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

ORF66 and o221 protein show 67% aa identity in 155aa overlap:

-189-

```

orf66  1  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
M F+  Q+ KALF L LFH+L+I +SNYLVO P I G HTTWGAFSFPFIFLATDLTV
o221   1  MNVFSQTQRYKALFWLSLFLHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

5  orf66  61  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
o221   61  RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASFMAAYA 120

10  orf66  121 IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155
+GQILD+ VEN+LR+ + WW+AP AST+ G+ DT
o221   121 LGQILDVHVENRLRQSRWWLAPTASTLFGNVSDT 155

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of *N.*

15 *meningitidis*:

```

                10      20      30      40      50      60
orf66.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
                |||
orf66a     MYAFTAAQQQKALFWLVLHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
                10      20      30      40      50      60

                70      80      90      100     110     120
orf66.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
                |||
20  orf66a  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
                70      80      90      100     110     120

                130     140     150
orf66.pep  IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT
                :|||
30  orf66a  LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIASF
                130     140     150     160     170     180

orf66a     VDYLFLKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX
                190     200     210     220

```

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

```

1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCTGGCTGGT
51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CCTTCCAAAT TTCCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
40 151 TTCACTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
201 GGCACGGCGG ATTATCTTTT GGGTCATGTT CCCCGCCCTT TTGCTTTCCCT
251 ACGTCTTTTC CGTTTGTGTC CACAACGGCA GTTGACGGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTCG
351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTTGTGTTT AACAAATTAC
45 401 GCCGTCTGAA AGCGTGTTGG GTTGCCCGCA CTGCATCAAC CGTCATCGGC
451 AACGCCTTAG ATACGTTGGT ATTTTTCGCC GTTGCTTCT ACGCAAGCAG
501 CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAACT CACCGTCTGC GGTCTGTTT TCCTGCCCGC CTACGGCGTG
601 ATTCTGAATC TGCTGACGAA AAACTGACG ACCCTGCAAA CCAAACAGGC
50 651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 264>:

```

1  MYAFTAAQQQ KALFWLVLFH ILIIAASNYL VQFPFQISGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA LGQILDIFVF NKLRLKAWW VAPTASTVIG
55 151 LNADTLVEFA VAFYASSDGF MAANWQGIAF VDYLFLKLTVC GLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*

```

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

```

                10      20      30      40      50      60
orf66a.pep  MYAFTAAQQQKALFWLVLHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
                |||
60  orf66-1  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV

```

-190-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf66a.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
	orf66-1	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
		70	80	90	100	110	120
10	orf66a.pep	LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
	orf66-1	IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
		130	140	150	160	170	180
15	orf66a.pep	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
	orf66-1	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
		190	200	210	220		
20		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N. gonorrhoeae*:

25	orf66.pep	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66ng	MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	120
30	orf66ng	RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA	120
	orf66.pep	IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT	155
	orf66ng	LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF	180

35 The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

```

1  ATGTACGCAT  TGACCGCCGC  ACAGCAACAG  AAGGCACTCT  TCCGGCTGGT
51  GCTTTTCCAT  ATCCTCATCA  TCGCCGCCAG  CAACTATCTG  GTGCAGTTCC
101 CCTTCCGAT   TTTCGGCATC  CACACCACTT  GGGCGCGGTT  TTCCTTTCCC
151 TTCATCTTCC  TCGCCACCGA  CCTGACCGTC  CGCATTTTCG  GTTCGCACTT
40  201  GCGCGCGCGG  ATTATCTTTT  GGGTGATGTT  CCCCGCCCTT  ttgCTTTcat
251  aCGTCTTTTC  CGTTTTGTTC  CACAACGGCA  GTTGACGGG   CTGGGCGCG
301  ctgTCCCAAT  TCAACACCTT  TGTCCGACGC  ATCGCGCTGG  CAAGTTTTC
351  CGCCTACGCG  CTCGGACAAA  TCCTTGATAT  TTTCGTATTC  GACAAATTAC
401  GCCGTCTGAA  AGCGTGGTGG  ATTGCCCGCG  CCGCATCAAC  CGTCATCGGC
45  451  AATGCACTGG  ACACGTTAGT  ATTTTTTGCC  GTTGCCTTTT  ACGCAAGCAG
501  CGATGAATTT  ATGGCGGCAA  ACTGGCAGGG  CATCGCTTTT  GTCGATTACC
551  TGTTCAAAC  TACCGTCTGC  ACCCTCTTCT  TCCTGCCCGC  CTACGGCGTG
601  ATACTGAATC  TGCTGACGAA  AAAACTGACG  GCCCTGCAAA  CCAAACAGCG
651  GCAAGACCGC  CCCGTGCCCT  CGCTGCAAAA  TCCGTAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 266>:

```

1  MYALTAAQQQ  KALFRLVLFH  ILIIAASNYL  VQFPFRIFGI  HTTWGAFSFP
51  FIFLATDLTV  RIFGSHLARR  IIFWVMFPAL  LLSYVFSVLF  HNGSWTGLGA
101  PSQFNTFVGR  IALASFAAYA  LGQILDIFVF  DKLRRLKAWW  IAPAASTVIG
151  NALDTLVFFA  VAFYASSDEF  MAANWQGI AF  VDYLFKLTVC  TLFFLPAYGV
55  201  ILNLLTKKLT  ALQTKQAQDR  PVPSLQNP*

```

An alternative annotated sequence is:

```

1  MYALTAAQQQ  KALFRLVLFH  ILIIAASNYL  VQFPFRIFGI  HTTWGAFSFP
51  FIFLATDLTV  RIFGSHLARR  IIFWVMFPAL  LLSYVFSVLF  HNGSWTGLGA
101  LSQFNTFVGR  IALASFAAYA  LGQILDIFVF  DKLRRLKAWW  IAPAASTVIG
60  151  NALDTLVFFA  VAFYASSDEF  MAANWQGI AF  VDYLFKLTVC  TLFFLPAYGV
201  ILNLLTKKLT  ALQTKQAQDR  PVPSLQNP*

```

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```

5  orf66-1.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV  60
   orf66ng      MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV  60

10 orf66-1.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
   orf66ng      RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120

15 orf66-1.pep  IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF 180
   orf66ng      LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180

   orf66-1.pep  VDYLFKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX  229
   orf66ng      VDYLFKLTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX  229

```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```

20  sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
   REGION (O221)
   >gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
   (U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
   hypothetical 25.3 kD protein in ftsY-nikA intergenic region [Escherichia coli]
   Length = 221
   Score = 273 bits (692), Expect = 5e-73
25  Identities = 132/203 (65%), Positives = 155/203 (76%)

   Query: 1  MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV  60
   Sbjct: 1  MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV  60

30  Query: 61  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
   Sbjct: 61  RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQFGALAHFNLFVARIATASF MAYA 120

35  Query: 121 LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
   Sbjct: 121 LGQILDVHVFENRLRQSRWWLAPTASTLFGNVSDTLAFFFI AFWRSPDAFMAEHWMEIAL 180

40  Query: 181 VDYLFKLTVCTLFFLPAYGVILN 203
   Sbjct: 181 VDYFK+ + +FFLP YGV+LN
   Sbjct: 181 VDYCFKVLISIVFFLPMYGVLLN 203

```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

```

50  1  ATGGTCATAA AATATACAAA TTGAATTTT GCGAAATTGT CGATAATTGC
   51  AATTTTGATG ATGTATTTCGT TTGAAGCGAA TGCAAyGCA GTmwrAATAT
  101  CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
   151  GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
  201  TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
   251  CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAACTTGCC
  301  CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTCCCTATG TCGGAACAGC
   351  CcTTTTAGCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
  401  GAGGCTACCA ATACGACCCC GAAACCGACA AATTGTAAA AGGCTACGAA
   451  TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA

```

-192-

501 TGGCTGCTAC GGC GTT GAT..

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF
 51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVKLA
 101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFKVKGYE
 151 YSNCLWYEDK RRINRTYGCY GVD..

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
 101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
 151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
 201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
 251 CCGCCAGCGT ATCCCGCGCC GCGTATTGG CGGGGGTCGG CAAACTTGCC
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
 15 351 CCTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
 401 GAGGCTACCA ATACGACCCC GAAACGACA AATTGCAAA GGTCTCAGGC
 451 TAA

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
 20 51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
 101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
 151 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
orf72.pep		MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF VPKNSKTYSS					
30 orf72a		MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf72.pep		DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVKLARLGA KFSTRA VPYVGTALLA					
35 orf72a		DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLARLGA KFSTRA VPYVGTALLA					
		70	80	90	100	110	120
		130	140	150	160	170	
40 orf72.pep		HDVYETFKEDI QARGYQYDP ETDKFKVKGYE YSNCLWYEDK RRINRTYGCY GVD					
orf72a		HDVYETFKEDI QARGYQYDP ETDKFAKVS GX					
		130	140	150			

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

45 1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
 101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
 151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
 201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
 50 251 CCGCCAGCGT ATCCCGCGCC GCGTATTGG CGGGGGTCGG CAAACTTGCC
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
 351 CCTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
 401 GAGGCTACCA ATACGACCCC GAAACGACA AATTGCAAA GGTCTCAGGC
 451 TAA

55 This encodes a protein having amino acid sequence <SEQ ID 272>:

-193-

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

```

5 ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

```

10 orf72a.pep      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANAVK ISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||||
10 orf72-1        10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANAVK ISETVSVD TGQGAKIHKF VPKNSKTYSS

15 orf72a.pep      70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA
    |||||
15 orf72-1        70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA

20 orf72a.pep      130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX
    |||||
20 orf72-1        130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX

```

Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from *N. gonorrhoeae*:

```

30 orf72.pep      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANAVX ISETVSVD TGQGAKIHKF VPKNSKTYSS
    || :|||
30 orf72ng        60
    MVTKHTNLNF AKLSIIAILM MYSFEANANAVK ISETLSVD TGQGAKVHKF VPKSSNIYSS

    orf72.pep      120
    DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA
    || :|||
    orf72ng        120
    DLTAKVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LVRQGAKF GTRAVPYV GTALLA

35 orf72.pep      173
    HDVYETFKED IQARGYQYDP ETDKFKVKG YEYSNCLWYED KRRINRTY GCGYVD
    ||||| :|||
35 orf72ng        180
    HDVYETFKED IQARGCRYDP ETDKFKVKG YEYANCLWYED ERRINRTY GCGYVDSS IMRLM

```

An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```

40 1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
51 VPKSSNIYSS DLTAKVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLV
101 RQGAKEGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKFKVKG YE
151 YANCLWYEDE RRINRTYGCY GVDSSIMRLM PDRSRFPEVK QLMESQMYRL
201 ARPFWNWRKE ELNKLSSLDW NNEVLNRCTF DWNGGGCAVN KGDDFRAGAS
45 251 FSLGRNPKYK EEMDAKKPEE ILSLKVDADP DKYIEATGYG GYSEKVEVAP
301 GTKVNMGPVT DRNGNPVQVA ATFGDAQGN TTADVQVIPR PDLTPASAEA
351 PHAQPLPEVS PAENPANPD PDENPGTRPN PEPDPLNPD ANPDTGQPG
401 TSPDSPAVID RPNGRHRKER KEGEDGGLSC DYFPEILACQ EMGKPSDRMF
451 HDISIPQVTD DKTWSSHNF LPSNGVCPQPK TFHVFGQYR ASYEPLCVFA
50 501 EKIRFAVLLA FIIMSAFVVF GSLGGE*

```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

```

55 1  ATGGTCACAA AACATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101 CTGAAACTCT TTCGGTTGAT ACCGACAAAG GCGCGAAAGT TCATAAGTTC
151 GTTCCTAAAT CAAGTAATAT TTATTCATCT GATTAAACAA AAGCGGTAGA
201 TTAAACGCAT ATCCCGACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGT CGGGGGTCGG CAACTTGTC
301 CGCCAAGGCG CGAAATTCGG CACAAGGGCG GTTCCTATG TCGGAACAGC
351 CCTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
60 401 GAGGCTGCCG ATACGATCCC GAAACCGACA AATT

```

This corresponds to the amino acid sequence <SEQ ID 276; ORF72ng-1>:

```

1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAHVHKKF
51  VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLSGVGLV
101 RQGAKEGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKF

```

5 ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

```

                                10      20      30      40      50      60
orf72ng-1.pe MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAHVHKKF VPKSSNIYSS
10 orf72-1      MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVDTGQGAHVHKKF VPKNSKTYSS
                                10      20      30      40      50      60

                                70      80      90      100     110     120
orf72ng-1.pe DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLSGVGLV RQGAKEGTRA VPYVGTALLA
15 orf72-1      DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLGAGVGLAR LGAKFSTRAPV VPYVGTALLA
                                70      80      90      100     110     120

                                130     140
orf72ng-1.pe HDVYETFKEDI QARGCRYDP ETDKF
20 orf72-1      HDVYETFKEDI QARGYQYDP ETDKFAKVSGX
                                130     140     150

```

Based on this analysis, including the presence of a putative leader sequence and transmembrane
 25 domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and
N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for
 raising antibodies.

Example 33

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 277>:

```

30 1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101 TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
151 GCTGACCGGT CTTTTATTGG CGGGCGCGGC AATGAGAAGC GCGGGGAAGG
201 TATCCGTTTA TCAGATGTTG TGGCCTATC..

```

35 This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:

```

1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLRQTG
51  LTGLLLAGAA MRSGBKSVSY QMLWPI..

```

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

```

40 1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101 TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCATACGGGG
151 CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
201 ATCCGTTTAT CAGATGTTGT GGCCTATCCG TTATACGGTG GCGGCTGTGT
251 GTCTGATGAG TCCGGGATTC GTATCCTCGG TGTTGGCGGT ATTGCTGCTG
45 301 CTGCCGTTTA AGGGAGGGGC AGTGTTGCAG GCAGGAGGTG CGGAAAATTT
351 TTTCAACATG AACCAATCGG GCAGAAAAGA GGGCTTTTCC CGCGATGACG
401 ATATTATCGA GGGAGAATAT ACGGTTGAAG AGCCTTACGG CGGCAATCGT
451 TCCCGAAACG CCATCGAACA CAAAAAAGAC GAATAA..

```

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

```

50 1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLRHTG
51  LSGLLLAGAA MRSGRVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101 LPFKGGAVLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEYPYGNR

```


Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of *N.*

5 *meningitidis:*

```

10      orf73.pep      MRFFGIGFLVLLFLEIMSI VVWADWLGGGWT LFLMAAGFAAGVLMRLQTGLTGLLLAGAA
                        10      20      30      40      50      60
      orf73a          MRFFGIGFLVLLFLEIMSI VVWADWLGGGWT LFLMAATFAAGVVMRLHTGLSGLLLAGAA
                        10      20      30      40      50      60

15      orf73.pep      MRSGGKVS VYQMLWPI
                        70
      orf73a          MRSGGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM

```

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

	1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
	51	GTCGATTGTG	TGGGTTGCCG	ATTGGTTGGG	CGGCGGTTGG	ACGCTGTTTC
20	101	TAATGGCGGC	AACCTTTGCC	GCCGCGGTGA	TGATGCTCAG	GCATACGGGG
	151	CTGTCCGGTC	TTTTATTGGC	GGGCGCGGCA	ATGAGAAGCG	GCGGGAGGGT
	201	ATCCCGTTAT	CANATGTTGT	GGCNTATCCG	TTATACGGTG	GCGGCGGTGT
	251	GTCNGATGAG	TCCGGGATTC	GTATCCTCGG	TGTNGGCGGT	ATTGCTGNTG
	301	CTNCCGTTTA	AGGGAGGTGC	AGTGTGTCAG	GCAGGAGGTG	CGGAAAATTT
25	351	TTTCAACATG	AACCANTCGG	GCAGAAAAGA	NGGCNTTTCC	CGCGATGACG
	401	ATATTATCGA	GGGGGAATAT	ACGGTTGAAG	ANCNTTACGG	CGGCANTCGT
	451	TTCCGAAACG	CNNTNGAACA	CAAAAAGAC	GAATAA	

This encodes a protein having amino acid sequence <SEQ ID 282>:

30

1	MRFFGIGFLV	LLFLEIMSIV	WVADWLGGGW	TLFLMAATFA	AGVVMLRHTG
51	LSGLLLAGAA	MRSGGRVSVY	XMLWXIRYTV	AAVCXMSPGF	VSSVXAVLLX
101	LPFKGGAVLQ	AGGAENFFNM	NXSGRXGXS	RDDDIIEGEY	TVEXPYGGXR
151	ERNAXEHKED	E*			

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

```

35      orf73a.pep      10      20      30      40      50      60
      MRFFGIGFLVLLFLEIMSI VVADWLGGGWTFLMAATFAAGVVM LRHTGLSGLLLAGAA
      |||||
      orf73-1      MRFFGIGFLVLLFLEIMSI VVADWLGGGWTFLMAAGFAAGVLM LRHTGLSGLLLAGAA
      10      20      30      40      50      60

40      orf73a.pep      70      80      90      100      110      120
      MRSGGRVSVYXMLWXIRYTVA AVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
      |||||
      orf73-1      MRSGGRVSVYQMLWP IRYTVA AVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
      70      80      90      100      110      120

45      orf73a.pep      130      140      150      160
      NXSGRKXGXS RDDDIIEGEYTVEXPYGGXRFRNAXEHKKDEX
      | ||||
      orf73-1      NQSGRKEGFSRDDDIIEGEYTV EEPYGGNRSRNAIEHKKDEX
      130      140      150      160

50

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N. gonorrhoeae*:

55 orf73.pep MRFFGIGFLVLLFLEIMSVVADWLGGGWTFLFMAAGFAAGVLMRLQTGLTGLLLAGAA 60
|||||

-196-

```

      orf73ng      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVLMRLHTGLSGLLLAGAA      60
      orf73.pep    MRSGBKVSQYQMLWPI                                                    76
      orf73ng      VKSSGKVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLLFPKGGAVLQAGGAENFFNM      120

```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```

      1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAAATTAT
      51  GTCGATTGTG TGGGTGCGCG ATTGGCTGGG CGGCGGTTGG AcgcTGTTC
     101  TAATGGCGGC AACCTTTGCC GCCGGTGTGC TGATGCTCAG GCATAcggGG
     151  CTGTCCGGTC TTTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
     201  ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
     251  GTCTGatgag tCcgGATTC GTATCCTccg tgttggCGGT ATTGCTGCTG
     301  CTGCcgttta aggGaggGgc agtgttgcag gcaggaggtg cggaaaATTT
     351  TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
     401  atattatcga cggagaatat acggttgaaa aacctgacgg cggcaatcgt
     451  tcccgaAAcg ccatcgaaca cgaaaAagac gaataA

```

This encodes a protein having amino acid sequence <SEQ ID 284>:

```

      1  MRFFGIGFLV LLFLEIMSIW VVADWLGGGW TLFLMAATFA AGVLMRLHTG
     51  LSGLLLAGAA VKSSGKVSQY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
     101  LFPKGGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGNR
     151  SRNAIEHEKD E*

```

ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```

      10      20      30      40      50      60
      orf73-1.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLHTGLSGLLLAGAA
      orf73ng      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVLMRLHTGLSGLLLAGAA
      10      20      30      40      50      60
      70      80      90      100     110     120
      orf73-1.pep MRSGBRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLLFPKGGAVLQAGGAENFFNM
      orf73ng      VKSSGKVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLLFPKGGAVLQAGGAENFFNM
      70      80      90      100     110     120
      130     140     150     160
      orf73-1.pep NQSGRKEGFSRDDDIIEGEYTVVEEPYGGNRSRNAIEHKKDEX
      orf73ng      NQSGRKEGFFHDDDIIEGEYTVVEKPDGGNRSRNAIEHEKDEX
      130     140     150     160

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 34

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 285>:

```

      1  ATGTTTGTTC TTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
     51  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
    101  TCGGCAATTT GCGCGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
    151  GCG..... .GCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
    201  CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
    251  GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
    301  GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
    351  ACTCGCCCGC CGCGTGCCTG AGGCCGGGTT TAAAGTCGTT CCCGTCGTGG
    401  GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
    451  GATTTTATT TCAACGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA

```

-197-

501 ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
 551 CGCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
 601 GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
 651 CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
 701 ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
 751 AAACACGAAG GCTTGTCGGA GTCCGCGCAA AACATCATGA AAATCCTCAC
 801 AGCCGAGCTG CCGACCAAAC AGCGGGCGGA GCTTGCTGCC AAAATCACGG
 851 GCGAGGGAAA GAAAGCTTTG TACGAT..

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

10 1 MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
 51 A....AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMV
 101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGAXAVMA ALSVAGVEGS
 151 DFYFNGFVPP KSGERRKLFA KWVRAAFPIV MFETPHRIGA ALADMAELFP
 201 ERRMLLAREI TKTFTFFLSG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE
 15 251 KHEGLSESAQ NIMKILTAEL PTKQAAELAA KITGEGKKAL YD..

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

20 1 ATGTTTCAGA AACATTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
 101 GCGCTTTGGC GGTATTGCAA AAGCGGGACA TCATCTGTGC CGAAGACACG
 151 CCGCTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACCTCGT
 201 CAGTGTGCGG GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCCGCT
 251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
 25 401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTCAACGG TTTGTACCG
 451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGCGGCG
 501 GTTTCCTATC GTCATGTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
 551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGGAA
 601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
 30 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG
 701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
 751 CAAAACATCA TGAAAACTCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
 801 GGAGCTTGCT GCCAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
 851 TGGCTCTGTC TTGGAAAAAC AAATAG

35 This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQKLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALS VAGVEG SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE
 40 201 ITKTFTFELS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N. meningitidis*:

50 orf75.pep MFVFQTAFXMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKAXXXXAEDTR
 orf75a MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTR
 55 orf75.pep VTAQLLSAYGIQKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR
 orf75a VTAQLLSAYGIQKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR
 orf75.pep RVREAGFKVVPVVGAXAVMAALS VAGVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIV

-198-

The complete length ORF75a nucleotide sequence <SEQ ID 289> is:

This encodes a protein having amino acid sequence <SEQ ID 290>:

ORF75a and ORF75-1 show 98.3% identity in 291 aa overlap:

BNSDOCID: <WO 9924578A2_1_>

-199-

	orf75-1	ATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD	
		190 200 210 220 230 240	
5	orf75a.pep	EKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALYDLALSWKNKX	
	orf75-1	EKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALYDLALSWKNKX	
		250 260 270 280 290	
10	<u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>		
	ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from <i>N. gonorrhoeae</i> :		
	orf75.pep	MFVFQTAFXMFKHLQKASDSVVG GTLYVVATPIGNLADITLRALAVLQKA----AEDTR	56
15	orf75ng	MSVFQTAFMFQKHLQKASDSVVG GTLYVVATPIGNLADITLRALAVLQKADIICAEDTR	60
	orf75.pep	VTAQLLSAYGIQGLVSVREHNERQMA DKIVGYLS DGMVVAQVSDAGTFAVCDPGAKLAR	116
20	orf75ng	VTAQLLSAYGIQGLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTFAVCDPGAKLAR	120
	orf75.pep	RVREAGFKVVPVVGASAVMAALSVA GVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFPV	176
	orf75ng	RVREAGFKVVPVVGASAVMAALSVA GVAESDFYFNGFVPPKSGERRKLF AKWVRAAFPV	180
25	orf75.pep	MFETPHRIGAAALADMAELFPERRMLL AREITKTFETFLSGTVGEIQTALSADGDQSRGEM	236
	orf75ng	MFETPHRIGATLADMAELFPERRMLL AREITKTFETFLSGTVGEIQTALAADGNQSRGEM	240
30	orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILT AE LPTKQAAELA AKITGEGKKALYD	288
	orf75ng	VLVLYPAQDEKHEGLSESAQNAMKILAE LPTKQAAELA AKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

35	1	MSVFQTAFM	FQKHLQKASD	SVVG GTLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTR	VTAQLLSAYG	IQGRLVSVRE	HNERQMA DKV	IGFLSDGLVV
	101	AQVSDAGTFA	VCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSVAGVAES
	151	DFYFNGFVPP	KSGERRKLEFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP
	201	ERRMLLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
40	251	KHEGLSESAQ	NAMKILAE	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
	301	*				

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

45	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTACAG	GCAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
	251	TCCTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
50	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAAATG	GCGGCGTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTGTACCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGC
	501	ATTTCCTGTC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
	551	CCGATATGGC	GGAATTGTTC	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
55	601	ATCACGAAAA	CGTTTGAAAC	GTTCTTAAGC	GGCACGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGTTG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751	CAAAATGCGA	TGAAAATCCT	TGCGGCCGAG	CTGCCGACCA	AGCAGGCGGC
	801	GGAGCTTGCC	GCCAAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
60	851	TGGCACTGTC	GTGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

-200-

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRLALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQGRVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFV VMFETPHRIG ATLADMAELF PERRLMLARE
 201 ITKTFTFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

		10	20	30	40	50	60
10	orf75-1.pep	MFQKHLQKASDSVVG	GTLYVVATPIGNLADIT	LRALAVLQKADIICAED	TRVTAQLLSAY		
	orf75ng-1	MFQKHLQKASDSVVG	GTLYVVATPIGNLADIT	LRALAVLQKADIICAED	TRVTAQLLSAY		
		10	20	30	40	50	60
15	orf75-1.pep	GIQGRVSVREHNERQ	MA DKIVGYLS DGMVVAQVSDAGT	PAVCDPGAKLARRVREAGFKV			
	orf75ng-1	GIQGRVSVREHNERQ	MA DKVIGFLSDGLVVAQVSDAGT	PAVCDPGAKLARRVREAGFKV			
		70	80	90	100	110	120
20	orf75-1.pep	VPVVGASAVMAALSVAGVEG	SDFYFNGFVPPKSGERRKLF	AKWVRAAFVVMFETPHRIG			
	orf75ng-1	VPVVGASAVMAALSVAGVAES	SDFYFNGFVPPKSGERRKLF	AKWVRAAFVVMFETPHRIG			
		130	140	150	160	170	180
25	orf75-1.pep	ATLADMAELFPERRLMLARE	ITKTFTFLSGTVGEIQTAL	SADGNQSRGEMVVLVLYPAQD			
	orf75ng-1	ATLADMAELFPERRLMLARE	ITKTFTFLSGTVGEIQTAL	AADGNQSRGEMVVLVLYPAQD			
		190	200	210	220	230	240
30	orf75-1.pep	EKHEGLSESAQNIMKILTAEL	PTKQAAELAAKITGEGKKA	LYDLALSWKNKX			
	orf75ng-1	EKHEGLSESAQNAMKILAAEL	PTKQAAELAAKITGEGKKA	LYDLALSWKNKX			
		250	260	270	280	290	
35	orf75-1.pep	EKHEGLSESAQNIMKILTAEL	PTKQAAELAAKITGEGKKA	LYDLALSWKNKX			
	orf75ng-1	EKHEGLSESAQNAMKILAAEL	PTKQAAELAAKITGEGKKA	LYDLALSWKNKX			
		250	260	270	280	290	

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)
 40 >gi|606086 (U18997) ORF_f286 [Escherichia coli]
 >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region [Escherichia coli] Length = 286
 Score = 218 bits (550), Expect = 3e-56
 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)
 45 Query: 4 KHLQKASDSVVG GTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
 K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
 Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLLQHFGIN 59
 50 Query: 64 GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGT PAVCDPGAKLARRVREAGFKVVPV 123
 RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
 Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119
 55 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIGATL 183
 G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
 Sbjct: 120 PGCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAIEAPRTLIFYESTHRLDSL 179
 60 Query: 184 ADMAELFPERR-LMLAREITKTFTFLSGTVGEIQTALAADGNQSRGEMVVLVLYPAQDEK 242
 D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMV++ +
 Sbjct: 180 EDIVAVLGESRYVVLARELTKTWTETIHGAPVGEELLAWKEDENRRKGEMVLIV-EGHKAQ 238
 65 Query: 243 HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKA LYDLAL 286
 E L A + +L AELP K+AA LAA+I G K ALY AL
 Sbjct: 239 EEDLPADALRTLALLQAEPLPKKAAALAAEIHGVKNALYKYAL 282

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 35

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

      1 ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
     51 TTTTGC GGCA GC.AAAGCAC CCGAAATCGA CCCGGCTTTG .....
           //
    651 ..... ..GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
    701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAGA AAACGGTGTC
    751 AAACCGTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

```

      1 MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
           //
    201 ..... ELVRNQLEQG LRQEKARLKI DALLEENGVK
    251 P*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

```

      1 ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
     51 TTTTGC GGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
    101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
    151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC
    201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
    251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
    301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAGACGAGCT
    351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
    401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
    451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
    501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC
    551 AGTTTGCCGC GATGAATCGG GGCGACGTTA CCCGCGATCC GGTCAAATTG
    601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
    651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
    701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAGA AAACGGTGTC
    751 AAACCGTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

```

    35      1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
           51 KPDGQAIRND AVRRILQTLV LKNRALKEGL DKDKDVQNRK KIAEASFYAE
           101 EYVRFLESE TVSEDELHKF YEQQIRMIKL QQVSFATEEE ARQAQQLLLK
           151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
           201 GERYYLFLKS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENGV
    40      251 KP*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) from strain A of *N. meningitidis*:

```

    45      orf76.pep      10      20      30
           MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL
           |||
    50      orf76a      10      20      30      40      50      60
           MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
           |||
           //
           70      80      90
  
```

-202-

```

orf76.pep                                XELVRNQLEQGLRQEKARLKIDALLEENGVPKX
                                         |||
orf76a      DVTRDPVKLGERYYLFLKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLKIDALEENGVPKX
              200      210      220      230      240      250

```

5 The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

```

1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51  TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAAC
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
15 451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCAGACCG CTGGCTTCGC
551 AGTTTCGAGC GATGAATCGG GCGCAGCTTA CCCGCGATCC GGTCAAATTG
601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTTGAGAC
20 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCA TTTTGAAGA AAACGGTGTC
751 AAACCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 300>:

```

1  MKQKKTA AAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51  KPDGQAIRND AVRRLQTLEV LKNRALKEGL DKDKDVQNRK KIAEASFYAE
25 101 EYVRFLE RSE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
201 GERYLFLKLS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALEENGVP
251 KP*

```

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

```

30 orf76a.pep      10      20      30      40      50      60
      MKQKKTA AAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
orf76-1      10      20      30      40      50      60
      MKQKKTA AAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND

35 orf76a.pep      70      80      90      100     110     120
      AVRRLQTLEV LKNRALKEGLDKDKDVQNRFKIAEASFYAE EYVRFLE RSE TVSESALRQF
orf76-1      70      80      90      100     110     120
      AVRRLQTLEV LKNRALKEGLDKDKDVQNRFKIAEASFYAE EYVRFLE RSE TVSEDELHKF

40 orf76a.pep      130     140     150     160     170     180
      YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
orf76-1      130     140     150     160     170     180
      YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP

45 orf76a.pep      190     200     210     220     230     240
      LASQFAAMNRGDVTRDPVKLGERYYLFLKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLK
orf76-1      190     200     210     220     230     240
      LASQFAAMNRGDVTRDPVKLGERYYLFLKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLK

50 orf76a.pep      250
      IDALEENGVPKX
orf76-1      250
      IDALLEENGVPKX
      250

```

60 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

-203-

```

      orf76.pep      MKQKKTAAAVIAAMLGFAAXKAPEIDPAL      30
                      |||
      orf76ng      MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND      60
                      //
5      orf76.pep      ELVRNQLEQGLRQEKKARLKDALLEENGVPK      251
                      |||
      orf76ng      VTRNPVKLGERYLLFKLGAVGKNPDAQPFELVRNQLEQGLRQEKKARLKDALLEENGVPK      251

```

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

```

10      1  ATGAAACAGA AAAAGACCGC TGCCGCGAGT ATTGCTGCAA TGTTGGCAGG
      51  TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
     101  TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
     151  AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTGCAAAC
     201  TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
     251  AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
     301  GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
     351  GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
     401  GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
     451  GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
     501  GTTCGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTcgc
     551  agtttgCCGG TATGAACCGT GGCAGCGTTA CCCGCAATCC GGTCAAATTG
     601  GCGCAACGCT ATTACCTGTT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
     651  GCGCAGCCTT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTGAGGC
     701  AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA Aaaccggtgtc
     751  AaacCGTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 302>:

```

      1  MKQKKTAAAV IAAMLGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
     51  RPDGQAIRND AVRRLQTLEV LKNRALKEGL DKDKDVQNRK KIAEASFYAE
    101  EYVRFLERSE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
    151  GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR GDVTRNPVKL
    201  GERYLLFKLG AVGKNPDAQP FELVRNQLEQ GLRQEKKARLK IDALLEENGVP
    251  KP*

```

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

```

      10      20      30      40      50      60
35      orf76-1.pep  MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND
      orf76ng      MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND
      10      20      30      40      50      60

      70      80      90      100     110     120
40      orf76-1.pep  AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLERSETVSEDELHKF
      orf76ng      AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLERSETVSESALRQF
      70      80      90      100     110     120

      130     140     150     160     170     180
45      orf76-1.pep  YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      orf76ng      YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      130     140     150     160     170     180

      190     200     210     220     230     240
50      orf76-1.pep  LASQFAAMNRGDVTRDPVKLGERYLLFKLSEVGKNPDAQPFELVRNQLEQGLRQEKKARLK
      orf76ng      LASQFAGMNRGDVTRNPVKLGERYLLFKLGAVGKNPDAQPFELVRNQLEQGLRQEKKARLK
      190     200     210     220     230     240

      250
60      orf76-1.pep  IDALLEENGVPKX
      orf76ng      IDALLEENGVPKX
      250

```

Furthermore, ORF76ng shows significant homology to a *B.subtilis* export protein precursor:

```

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
33K lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
[Bacillus subtilis]
5 >gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
>gi|2633331|gnl|PID|e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

10 Query: 70 VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAEYVRFLESETVSE 114
VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
Sbjct: 53 VLTQLVQEKVLDKKYKVSDEIDNKLKEYKTQLGDQYTALEKQYGYKDYLLKEQVKYELLTQ 112

15 Query: 115 SA-----LRQFYERQIRMIKLQOVSFATEEEARQAQQLLKGLSFEGLMKRYPN 163
A +++++E I+ + A ++ A + ++ L KG FE L K Y
Sbjct: 113 KAAKDNIVTDADIKEYWGLKGRASHILVADKKTAEEVEKKLKGKGFEDLAKEYST 172

Query: 164 DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLGERYYLFKLSEVGKNPDA 218
D A G F Q+ E + + G+V+ DPVK Y++ K +E D
20 Sbjct: 173 DSSASKGGDLGWFAKEGQMDETFSKAAFKLKTGEVS-DPVKTQYGYHIIKKTEERGKYDD 231

Query: 219 QPFELVRNQLEQGLRQKA 237
EL LEQ L A
25 Sbjct: 232 MKKELKSEVLEQKLNDA 250

```

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

35 Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

```

1 ATGAAAAAAT CTTTCCTTAC GCTTGTCTCTG TATTCGTCTT TACTTACCGC
51 CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG
101 GCAAAAATTG CGGAAACGTT TGCGCTGACA TTTGTGATTG CTGCGCTGTA
151 TCTGTTTGGC CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
201 CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
251 ATGACG.... //

1201 ..... CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
45 1251 ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
1301 ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCCGTAGT
1351 GTTGACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
1401 CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCAACGTT CCTGATTAC
1451 ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTGCGGAAG GCTCGGTAAC
50 1501 GGGCAACCTG ATTACGGGTG ATGCAAGCAG CTTGAACATT CGCGACGGCA
1551 AGGCGGAATA TGTTTATCCG CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

```

1 MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAAKIAETF ALTFVIAALY

```

-205-

51 LFARNKVTRL LIAVFFAFSI IANNVHYADY QSWMT.....
 401 ...QTVFEQL QKTPDGNWLF AYTS DHGQYV RQDIYNQGT V QPDSYLVPLV
 451 LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLG YDMP VSGCREGSVT
 501 GNLTGDAGS LNIRDGKA EY VYPQ*

Further work revealed the complete nucleotide sequence <SEQ ID 305>:

1 ATGAAAAAAT CTTTCCTTAC GCTTGTCTG TATTCGTCTT TACTTACCGC
 51 CAGCGAAATT GCCTATCGCT TTGTATTTGG GATTGAAACC TTACCGGCGG
 101 CAAAAATTGC GGAAACGTTT GCGCTGACAT TTGTGATTGC TCGCGTGTAT
 151 CTGTTTGCGC GTTATAAGGT GACGCGTTTG TTGATTGCGG TGTTTTTTGC
 201 GTTCAGCATT ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
 251 TGACGGGCAT CAATTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
 301 AGCGCGGGTG CGTCGATGTT GGATAAGTTG TGGCTGCCTG TGTGTGGGG
 351 CGTGTGGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
 401 CGCATTTTTC TGCCGATATA CTGTTTGCC TCCATATGCT GATGATTTTC
 451 GTGCGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTTCG CCAAACCGAC
 501 ATACAGCCGC ATCAAAGCCA ATTATTTTCA CTTCGGTTAT TTTGTCGGAC
 551 GCGTGTGGCC GTATCAGTTG TTTGATTAA GCAGGATTCC CGCCTTTAAG
 601 CAGCCTGCTC CAAGCAAAT CCGGCAGGGC AGTGTTCAAA ATATCGTCCT
 651 GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAGCTG TTTGGCTACG
 701 GACGCGAAAC TTCGCGGTTT TTAACCCGGC TGTCGCAAGC CGATTTTAA
 751 CCGATTGTGA AACAAAGTTA TTCCGACGGC TTTATGACTG CAGTGTCCT
 801 GCCCAGTTT TTCAATGCGA TACCGCACGC CAACGGCTTG GAACAAATCA
 851 GCGGCGGCGA TACCAATATG TTCCGCTCG CCAAAGAGCA GGGCTATGAA
 901 ACGTATTTT ACAGCGCGCA GCGGGAAC GAGATGGCGA TTTTGAACCT
 951 AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTGGCT
 1001 ACGGCAACGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTTCGAC
 1051 AAAATCAATT TGCAGCAGGG CAAGCATTTT ATCGTGTGAC ACCAACGCGG
 1101 TTCGCACGCC CCATACGGCG CATTGTTGCA GCCTCAAGAT AAAGTATTCG
 1151 GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
 1201 CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
 1251 CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTT CGCCAAGATA
 1301 TCTACAATCA AGGCACGGTG CAGCCGACA GCTATCTCGT GCCGCTAGTG
 1351 TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
 1401 GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTCACA
 1451 CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACG
 1501 GGCAACCTGA TTACGGGTGA TGCAGGCAGC TTGAACATTC GCGACGGCAA
 1551 GCGGGAATAT GTTTATCCGC AATGA

This corresponds to the amino acid sequence <SEQ ID 306; ORF81-1>:

40 1 MKKSFLTLVL YSSLLTASEI AYRFVFGIET LPAAKIAETF ALTFVIAALY
 51 LFARYKVTRL LIAVFFAFSI IANNVHYAVY QSWMTGINYW LMLKEVTEVG
 101 SAGASMLDKL WLPVLWGVLE VMLFCSLAKF RKRTHFSADI LFAFLMLMIF
 151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSPAPFK
 45 201 QPAPSKIGQG SVQNIVLIMG ESESAHLKL FGYGRETSPP LTRLSQADF
 251 PIVKQSYSAG FMTAVSLPSF FNAIPHANGL EQISGGDTNM FRLAKEQGYE
 301 TYFYSAQAEN EMAILNLIGK KWIDHLIQPT QLGYNGDNM PDEKLLPLFD
 351 KINLQQGKHF IVLHQRGSHA PYGALLQPD KVFGEADIVD KYDNTIHKTD
 401 QMIQTVFEQL QKQPDGNWLF AYTS DHGQYV RQDIYNQGT V QPDSYLVPLV
 451 LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLG YDMP VSGCREGSVT
 50 501 GNLTGDAGS LNIRDGKA EY VYPQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

55 orf81.pep MKKSFLTLVL YSSLLTASEI AYRFVFGIET LPAAKIAETF ALTFVIAALY LFARNKVTRL
 orf81a MKKSLFVLFL YSSLLTASEI AYRFVFGIET LPAAKMAETF ALTFVIAALY LFARYKATRL
 60 10 20 30 40 50 60
 70 80

-206-

```

orf81.pep  LIAVFFAFSIIANNVHYADYQSWMT
            |||||
orf81a     LIAVFFAFSIIANNVHYAVYQSWITGINYWLMKEITEVGGAGASMLDKLWLPALWGVLE
            70      80      90      100      110      120

5          //

orf81.pep                                     QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD
            |||||
orf81a     IPHANGLEQISGGDIVDKYDNTIHKTDQMIQTVFEQLQKTPDGNWLFAYTSDHGQYVRQD
            280      290      300      310      320      330

10         150      160      170      180      190      200
orf81.pep  IYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG
            |||||
orf81a     IYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG
            340      350      360      370      380      390

15         210      220      230
orf81.pep  CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
            |||||
orf81a     CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
            400      410      420

```

The complete length ORF81a nucleotide sequence <SEQ ID 307> is:

```

25      1  ATGAAAAAAT CCCTTTTTCGT TCTCTTTCTG TATTCGTCCC TACTTACTGC
      51  CAGCGAAATT GCTTATCGCT TTGTATTCGG AATTGAAACC TTACCGGCTG
     101  CAAAAATGGC AGAACGTTT GCGCTGACAT TTGTGATTGC TGCGCTGTAT
     151  CTGTTTGCGC GTTATAAGGC AACGCGTTG TTGATTGCGG TGTTTTTCGC
     201  GTTCAGCATT ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
     251  TAACGGGCAT TAATTATTGG CTGATGCTGA AAGAGATTAC CGAAGTTGGC
     301  GGCGCAGGGG CGTCGATGTT GGATAAGTTG TGGCTGCCTG CGTTGTGGGG
     351  CGTGTTGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
     401  CGCATTTTTT TCCCGATATA CTGTTTGCTT TCCTAATGCT GATGATTTTC
     451  GTGCGTTTCG TCGACACGAA ACAAGAACAC GGTATTTTCG CCAAACCGAC
     501  ATACAGCCGC ATCAAGCCA ATTATTTTCT CTTGCGTTAT TTTGTGCGAC
     551  GCGTGTTGCC GTTACAGTTG TTTGATTTAA GCAAGATTCC TGTGTTCAAA
     601  CAGCCTGCTC CAAGCAGAA CCGGCAAGGC AGTATTCAAA ATATCGTCCT
     651  GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAATTG TTTGGCTACG
     701  GGCGCGAAAC TTCGCCGTTT TTGACCCAGC TTTGCGCAAG CGATTTTAAAG
     751  CCGATTGTGA AACAAAGTTA TTCCGCAGGC TTTATGACGG CAGTATCCCT
     801  GCCCAGTTTC TTTAACGTCA TACCGCATGC CAACGCTTG GAACAAATCA
     851  GCGGCGGCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAACCGGAC
     901  CAAATGATT CAAACGTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
     951  CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTT CGCCAAGATA
    1001  TCTACAATCA AGGCACGGTG CAGCCCAGCA GCTATCTCGT GCCGCTGGTG
    1051  TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
    1101  GCCTTGCGAG ATTGCCCTCC ATCAGCAGCT TTCAACGTTT CTGATTACACA
    1151  CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACG
    1201  GGCAACCTGA TTACGGGTGA TGCAGGCAGC TTGAACATTC GCGACGGCAA
    1251  GGCGGAATAT GTTTATCCGC AATGA

```

50 This encodes a protein having amino acid sequence <SEQ ID 308>:

```

      1  MKKSLFVLFL YSSLLTASEI AYRFVFGIET LPAAKMAETF ALTFVIAALY
     51  LFARYKATRL LIAVFFAFSI IANNVHYAVY QSWITGINYW LMLKEITEVG
    101  GAGASMLDKL WLPALWGVLE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
    151  VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLISKIPVEK
    201  QPAPSRIGQG SIQNIIVLIMG ESESAHLKL FGYGRETSPP LTQLSQADFK
    251  PIVKQSYSAG FMTAVSLPSF FNVIPHANG EQISGGDIVD KYDNTIHKTD
    301  QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
    351  LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
    401  GNLITGDAGS LNIRDGKAEY VYPQ*

```

60 ORF81a and ORF81-1 show 77.9% identity in 524 aa overlap:

```

      10      20      30      40      50      60
orf81a.pep MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL
            |||:::|
orf81-1     MKKSFLTLLVYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL
            10      20      30      40      50      60

```

-207-

		70	80	90	100	110	120
	orf81a.pep	LIAVFFAFSIIANNVHYAVYQSWITGINYWMLKEITEVGGAGASMLDKLWLPALWGVLE					
5	orf81-1	LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPVLWGVLE					
		70	80	90	100	110	120
	orf81a.pep	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY					
10	orf81-1	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY					
		130	140	150	160	170	180
	orf81a.pep	FVGRVLPYQLFDLSKIPVFKQPAPSRIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPF					
15	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFYGRGRETSPF					
		190	200	210	220	230	240
	orf81a.pep	LTQLSQADFKPIVKQSYSAGFMTAVSLPSFFNVIPHANGLEQISGGD-----					
20	orf81-1	LTRLNQADFKPIVKQSYSAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFRLAKEQGYE					
		250	260	270	280	290	300
25	orf81a.pep	-----					
	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF					
30		310	320	330	340	350	360
	orf81a.pep	-----		290	300	310	320
	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTQMIQTVFEQLQKQPDGNWLF					
35		370	380	390	400	410	420
	orf81a.pep	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
40	orf81-1	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
	orf81a.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAHYVYPQX					
45	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAHYVYPQX					
		490	500	510	520		

50 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from *N. gonorrhoeae* of the N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

	orf81.pep	MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALTFFVIAALYLFARNKVTRL	60
55	orf81ng	MKKSFLVFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFFMIAALYLFARYKASRL	60
	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT	85
	orf81ng	LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPALWGVAE	120
60	orf81.pep	QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD	433
	orf81ng	ALLQPDQKVFGEADIVDKYDNTIHKTQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD	433
65	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81ng	IYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493

-208-

```
orf81.pep    CREGSVTGNLITGDAGSLNIRDGKAEYVYPQ    524
|||||:|||||
orf81ng      CREGSVTGNLITGDAGSLNIRNGKAEYVYPQ    524
```

The complete length ORF81ng nucleotide sequence <SEQ ID 309> is:

```
5      1  ATGAAAAAAT CCCTTTTCGT TCTCTTCTG TATTCATCCC TACTTACCGC
      51  CAGCGAAATC GCCTATCGCT TTGTATTCGG AATTGAAACC TTACCGGCTG
     101  CAAAAATGGC GGAAACGTTT GCGCTGACAT TTATGATTGC TGCGCTGTAT
     151  CTGTTTGCGC GTTATAAGGC TTCGCGGCTG CTGATTGCGG TGTTTTTTCGC
     201  GTTCAGCATG ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
     251  TGACGGGTAT TAACTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
     301  AGCGCGGGCG CGTCGATGTT GGATAAGTTG TGGCTGCCTG CTTTGTGGGG
     351  CGTGGCGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
     401  CGCATTTTTT TGCCGATATA CTGTTTGCCT TCCTAATGCT GATGATTTTC
     451  GTGCGTTTCG TCGACACGAA ACAAGAGCAC GGTATTTTCG CCAAACCGAC
     501  ATACAGCCGC ATCAAAGCCA ATTATTTTCAG CTTCGGTTAT TTTGTGGGG
     551  GCGTGTTGCC GTATCAGTTG TTTGATTTAA GCAAGATCCC TGTGTTCAAA
     601  CAGCCTGCTC CAAGCAAAAT CGGGCAAGGC AGTATTCAAA ATATCGTCCT
     651  GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAATTG TTTGGTTACG
     701  GCGCGGAAAC TTCGCGTTT TTAACCCGGC TGTGCAAGC CGATTGTGAG
     751  CCGATTGTGA AACAAAGTTA TTCCGCAGGC TTTATGACGG CAGTATCCCT
     801  GCCAGTTTTC TTTAACGTCA TACCGCACGC CAACGGCTTG GAACAAATCA
     851  GCGGCGGCGA TACCAATATG TTCCGCCTCG CCAAAGAGCA GGGCTATGAA
     901  ACGTATTTTT ACAGTGCCCA GGCTGAAAAC CAAATGGCAA TTTTGAACCT
     951  AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTGGCT
    1001  ACGGCAACGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
    1051  AAAATCAATT TGCAGCAGGG CAGGCATTTT ATCGTGTTGC ACCAACGCGG
    1101  TTGCGACGCC CCATACGGCG CATTGTTGCA GCCTCAAGAT AAAGTATTCG
    1151  GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
    1201  CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
    1251  CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTG CGCCAAGATA
    1301  TCTACAATCA AGGCACGGTG CAGCCCGACA GCTATATTGT GCCTCTGGTT
    1351  TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
    1401  GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTACA
    1451  CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACA
    1501  GGCAACCTGA TTACGGGCGA TGCAGGCAGC TTGAACATTC GCAACGGCAA
    1551  GCGGAATAT GTTTATCCGC AATAA
```

This encodes a protein having amino acid sequence <SEQ ID 310>:

```
      1  MKKSLFVLFL YSSLLTASEI AYRFVFGIET LPAAKMAETF ALTFMIAALY
     51  LFARYKASRL LIAVFFAFSM IANNVHYAVY QSWMTGINYW LMLKEVTEVG
    101  SAGASMLDKL WLPALWGVAE VMLFCSLAKF RRKTHFSADI LEAFMLLMIF
    151  VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSKIPVFK
    201  QPAPSKIGQG SIQIVLIMG EESAHLKL FGYGRETSPF LTRLSQADFK
    251  PIVKQSYSAG FMTAVSLPSF FNVIPHANGL EQISGGDTNM FRLAKEQGYE
    301  TYFYSAQAEI QMAILNLIGK KWIDHLIPT QLYGNGDNM PDEKLLPLFD
    351  KINLQQRHF IVLHQRGSHA PYGALLQPD KVFGEADIVD KYDNTIHKTD
    401  QMIQTVFEQL QKQPDGNWLF AYTSDHQYV RQDIYNQGTV QPDSYIVPLV
    451  LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYPDMP VSGCREGSVT
    501  GNLITGDAGS LNIRNGKAEY VYPQ*
```

ORF81ng and ORF81-1 show 96.4% identity in 524 aa overlap:

```
50      10      20      30      40      50      60
orf81ng-1.pep MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFMIAALYLFARYKASRL
           |||:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf81-1      MKKSFLTFLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL
           10      20      30      40      50      60

55      70      80      90     100     110     120
orf81ng-1.pep LIAVFFAFSMIANNVHYAVYQSWMTGINYWLMMLKEVTEVGSGASMLDKLWLPALWGVAE
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf81-1      LIAVFFAFSIIANNVHYAVYQSWMTGINYWLMMLKEVTEVGSGASMLDKLWLPVLWGVLE
           70      80      90     100     110     120

60      130     140     150     160     170     180
orf81ng-1.pep VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf81-1      VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
```

-209-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf81ng-1.pep	FVGRVLPYQLFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPF					
	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFYGRGRETSPF					
		190	200	210	220	230	240
10	orf81ng-1.pep	LTRLAQADFKPIVKQSYSGFMTAVSLPSFFNVIPHANGLEQISGGDTNMFLAKEQGYE					
	orf81-1	LTRLAQADFKPIVKQSYSGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFLAKEQGYE					
		250	260	270	280	290	300
15	orf81ng-1.pep	TYFYSAQAENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGRHF					
	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGRHF					
		310	320	330	340	350	360
20	orf81ng-1.pep	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVEQLQKQPDGNWLF					
	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVEQLQKQPDGNWLF					
25		370	380	390	400	410	420
	orf81ng-1.pep	AYTSDHGQYVRQDIYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
	orf81-1	AYTSDHGQYVRQDIYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
30	orf81ng-1.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRNGKAEYVYPQX					
	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRNGKAEYVYPQX					
		490	500	510	520		

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

40	gi1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)
45	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIAVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHVMQRLFFVLTILVVKRISSPLRLLLVAAPFVL-LTAADMSISLY- 86
50	Query: 82 SWMT-----GINYWMLKEVTEVGSAGASMLDKLWLPALWGVAEVMFLFCSLAKFRKRT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV---AKMLG-MYSPYLCAFAFLSLLFLAVIICYDV 141
55	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPKPTYSRIKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKKVTGILLIIVISGSLFSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201
60	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPFL 241 +Q L . + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVPYFQL----SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257
65	Query: 242 TRLSQADFKPIVKQSYSGFMTAVSLP---SFFNVIPHANGLEQISGGDTNMFLAKEQG 298 +Q + Q+ S TA+S+P + +V+ H I N+ +A + G Sbjct: 258 E--AQRKQIKLFNQAISGAPYTALSVPLSLTADSVLSH-----DIHNPNDNIINMANQAG 310
70	Query: 299 YETIFYSAQA--ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLLPPLSLQALQQ 359
	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPDQKVFGEADIVDK-YDNTIHKTDQMIQTVEQLQK 412 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+ Sbjct: 360 NTQOKKLIVLHLNGSHEPACSAYPQSSAVFQPDQDDQACDYDNSIHYTDSLLGQVFELLK- 418

Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQG--TVQPDSYIVPL-VLYSP 454
 D Y +DHG ++++Y G +Y VP+ + YSP
 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVPMPFIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

```

1   ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGCG GCACACTGAC
51  CCGCATACTC GCCCaCGGCG GCGGCAAACG CTTTGCCGTC GAACAAGAAC
101 TCGTCGCCGC ATCGTCCCGC GCCGCCGTCA AAGAAATGGA TTGTGCCGCC
15  yTAAAGGAC GCAAAGCCGC CyTTTACGTC TCCGTTATGG GCGACCAAGG
201 TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
251 GCGGCTACCA CAACAACCCC GAAAGTGCCA CCAATACAG CTACCCCGCC
301 TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
20  TTCCACATCG CTTTGAACG CCCCCGCCGC CGyCyTGACG AAAACAGCG
401 GACGCAAAGG CGAACGcTCC GCCGGACTGT CCGTCAACGG CACGGGCGAC
451 TACCGCAACG AAACCCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
501 CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCgC
551 CCGrATACGC CGACACCGAC GTATTCTGTA CCGTCGACGT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

```

1   ..TLLLFIPVL TXCGTLTGIL AHGGGKRFV EQELVAASSR AAVKEMDLA
51  LKGRKAAXYV SVMGDQSGN ISGGRYSIDA LIRGGYHNNP ESATQYSYPA
101 YDTTATTKSD ALSSVTTSTS LLNAPAAXLT KNSGRKGRS AGLSVNGTGD
151 YRNETLLANP RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFVTVDV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

```

1   ATGAAACCC TGCTCCTCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
51  ACTGACCGGC ATACCCGCCC ACGGCGGCGG CAAACGCTTT GCCGTGCAAC
101 AAGAACTCGT CGCCGCATCG TCCCGCGCCG CCGTCAAAGA AATGGATTTG
151 TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
35  CCAAGGTTTC GGCAACATAA GCGGCGGACG CTA CTCTATC GACGCACTGA
251 TACGCGGCGG CTACCACAAC AACCCTGAAA GTGCCACCCA ATACAGCTAC
301 CCGCCTACG ACCTACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT
351 AACCACTTCC ACATCGCTTT TGAACGCCCC CGCCGCGGCC CTGACGAAAA
401 ACAGCGGACG CAAAGCGGAA CGCTCCGCCG GACTGTCCGT CAACGGCACG
451 GGCGACTACC GCAACGAAAC CCTGCTCGCC AACCCTCGCG ACGTTTCCTT
40  CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
551 TACCGCCCGA ATACGCGGAC ACCGACGTAT TCGTAACCGT CGACGTATTC
601 GGCACCGTCC GCAGCCGTAC CGAAGTGCAC CTCTACAACG CCGAAACCTT
651 TAAAGCCCAA ACCAAGCTCG AATATTTTCG CCGTGACCGC GACAGCCGGA
701 AACTGCTGAT TACCCTAAA ACCGCGCGCT ACGAATCCCA ATACCAAGAA
45  CAATACGCCC TTTGGACCGG CCCTTACAAA GTCAGCAAAA CCGTCAAAGC
801 CTCAGACCGC CTGATGGTCG ATTCTCCGA CATTACCCCC TACGGCGACA
851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCC
901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

```

1   MKTLLLLIPL VLTACGLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
51  SALKGRKAAL YVSMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY
101 PAYDTTATTK SDALSSVTT TSLNAPAAA LTKNSGRKGE RSAGLSVNGT
151 GDYRNETLLA NPRDVSFLTN LIQTVFYLRG IEVVPPEYAD TDVFVTVDFV
201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYQE
55  QYALWTGPYK VSKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNKKPK
```


Homology with a predicted ORF from *N.meningitidis* (strain A)

5 *meningitidis:*

		10	20	30	40	50	
	orf83.pep	<u>TLLLFIPVLVLT</u> XCGLTGTGILAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAX					
	orf83a	<u>MKTLLXLIPLVLT</u> ACGLTGTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL					
10		10	20	30	40	50	60
		60	70	80	90	100	110
	orf83.pep	YVSVMGDQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS					
	orf83a	YVSVMGDQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS					
15		70	80	90	100	110	120
		120	130	140	150	160	170
	orf83.pep	TSLLNAPAAXLTKNSGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNLNIQTVFYLRG					
	orf83a	TSLLNAPAAALTKNSGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNLNIQTVFYLRG					
20		130	140	150	160	170	180
		180	190				
	orf83.pep	IEVVPPXYADTDVFVTVDV					
	orf83a	IEVVPEYADTDVFVTVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDRSKLLIAPK					
25		190	200	210	220	230	240

30	1	ATGAAACCC	TGCTCNCCT	CATCCCCCTC	GTCTCTACAG	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCCC	ACGGCGGCGG	CAAACGCTTT	GCCGTGCAAG
	101	AAGAACTCGT	CGCCGCATCG	TCCCGCGCGC	CCGTCAAAGA	AATGGACTTG
	151	TCCGCCCTGA	AAGGACGCAA	AGCCGCCCTT	TACGTCTCCG	TTATGGGCGA
35	201	CCAAGTTTCG	GGCAACATAA	GCGGCGGACG	CTACTCTATC	GACGCACTGA
	251	TACGCGGCGG	CTACCAACAAC	AACCCCGAAA	GTGCCACCCA	ATACAGCTAC
	301	CCCGCCTACG	ACACTACCGC	CACCACCAA	TCCGACGCGC	CTCCAGCGC
	351	AACCACCTTC	ACATCGCTTT	TGAACGCCCC	CGCCGCCGCC	TGACGAAAA
40	401	ACAGCGGACG	CAAAGGCGAA	CGCTCCGCCG	GACTGTCCGT	CAACGGCACG
	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCCCGG	ACGTTTCCTT
	501	CCTGACCAAC	CTCATCCAAA	CCGTCTTCTA	CCTCGCGCGC	ATCGAAGTCG
	551	TACCGCCCGA	ATACGCCGAC	ACCGACGTAT	TCGTAACCGT	CGACGTATTC
45	601	GGCACCCTCC	GCAGCCGCAC	CGAACTGCAC	CTCTACAACG	CCGAAACCGT
	651	TAAAGCCCAA	AGCAAGCTCG	AATATTTTCG	CGTTGACCGC	GACAGCCGGA
	701	AACTGCTGAT	TGCCCTTAAA	ACCGCCGCCT	ACGAATCCCA	ATACCAAGAA
	751	CAATACGCCC	TCTGGATGGG	ACCTTACAGC	GTGGGCAAAA	CCGTCAAAGC
45	801	CTCAGACCGC	CTGATGGTCG	ATTTCTCCGA	CATCACCCCC	TACGGCGACA
	851	CAACCGCCCA	AAACCGTCCC	GACTTCAAAC	AAAACAACGG	TAAAAAACCC
	901	GATGTCGGCA	ACGAAGTCAT	CCGCCCGCGC	AAAGGAGGAT	AA

50	1	MKTLLXLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSVMGDQGS	GNISGGRYSI	DALIRGGYHN	NPESATQYSY
	101	PAYDTTATTK	SDALSSVTTT	TSLLNAPAAA	LTKNRSGRKE	RSAGLSVNGT
	151	GDYRNETLLA	NPRDVSFLT	LIQTVFYLRG	IEVVPPEYAD	TDVFTVDVDF
	201	GTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLLIAPK	TAAYESQYQE
55	251	QYALWMGPYS	VGKTVKASDR	LMVDFSDITP	YGDTTAQNRP	DFKQNNKKKP
	301	DVGNEVIRRR	KGG*			

orf83a.pep 10 20 30 40 50 60
MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL

10

15

20

25

30

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N. gonorrhoeae*:

40

45

50

orf83.pep	IEVVPXPYADTDVFVTVDDV	197
orf83na	IEVVPPEYADTDVFVTVDFGTVRSRTELHLYNAETLKAOTKLEYFAVDRDSRKLIIAPK	240

The complete length ORF83ng nucleotide sequence <SEO ID 317> is:

55

50

65

-213-

5
 651 TAAAGCCCAA ACCAAGCTCG AATATTTTCGC CGTCGACCGC GACAGCCGGA
 701 AACTGCTGAT TGCCCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA
 751 CAATACGCC CTTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC
 801 CTCAGACCGC CTGATGGTTCG ATTTCTCCGA CATCACCCCG TACGGCGACA
 851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCCC
 901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10
 1 MKTLLLLLIPL VLTACGTLTG IPAHHGGGKRF AVEQELVAAS SRAAVKEMDL
 51 SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPDSATRYSY
 101 PAYDTTATTK SDALSGVTTS TSLLNAPAAA LTKNNGRKGE RSAGLSVNGT
 151 GDYRNETLLA NPRDVSFLTNI LIQTVFYLRG IEVVPPEYAD TDVFTVDFV
 201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE
 251 QYALWMPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNGKNP
 301 DVGNEVIRRR KGG*

15 ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

		10	20	30	40	50	60
orf83-1.pep	MKTLLLLLIPLVLTACGTLTGIPAHHGGGKRF	AVEQELVAASSRAAVKEMDL	SALKGRKAAL				
orf83ng	MKTLLLLLIPLVLTACGTLTGIPAHHGGGKRF	AVEQELVAASSRAAVKEMDL	SALKGRKAAL				
		10	20	30	40	50	60
orf83-1.pep	YVSVMGDQGS	GNISGGRYSIDALIRGGYHNPESATQSY	PAYDTTATTKSDALSSVTTS				
orf83ng	YVSVMGDQGS	GNISGGRYSIDALIRGGYHNPDSATRYSPAYDTTATTKSDALSGVTTS					
		70	80	90	100	110	120
orf83-1.pep	TSLLNAPAAALTKNSGRK	GERSAGLSVNGTG	DYRNETLLANPRDVSFLTNI	LIQTVFYLRG			
orf83ng	TSLLNAPAAALTKNNGRK	GERSAGLSVNGTG	DYRNETLLANPRDVSFLTNI	LIQTVFYLRG			
		130	140	150	160	170	180
orf83-1.pep	IEVVPPEYADTDVFTVDFGT	VRSRTELHLYNAETLKAQTKLEYFAVDR	DSRKLLITPK				
orf83ng	IEVVPPEYADTDVFTVDFGT	VRSRTELHLYNAETLKAQTKLEYFAVDR	DSRKLLIAPK				
		190	200	210	220	230	240
orf83-1.pep	TAAYESQYQEQYALW	TGPKVSKTVKASDRLMVDFSDITPYGDTTAQNRP	PDFKQNNGKKP				
orf83ng	TAAYESQYQEQYALW	MPYSVGKTVKASDRLMVDFSDITPYGDTTAQNRP	PDFKQNNGKNP				
		250	260	270	280	290	300
orf83-1.pep	DVGNEVIRRRKGGX						
orf83ng	DVGNEVIRRRKGGX						
		310					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

      1  ATGGCAGAGA TCTGTTTGTAT AACCGGCACG CCCGGTTCAG GGAAAACATT
5      51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
      101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAGGCTT GAAAATACCG
      151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
      201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
      251  TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
10     301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
      351  ACATCAGGGC ATTGATATAT TTGTTTTGAC TCAAGGTCCT AAGCTTCTAG
      401  ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACAT CGCTTCAAAC
      451  AAGATGGGTA TGCGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
      501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
15     551  AAGTTTATGA CTTGTATsrr TmmGCGGAAG TTCATACCGT AAATAAGGTC
      601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
      651  CGTGTTTGTC GGCCTGTCCT ATAAAATGTT GagCaGTTAC GGAAAAAAAC
      701  aGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
      751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
20     801  AGATATGTTT GTTCCGACAT TGTCCGaaAA ACCCGrAAGC AAGCgaTtT
      851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
      901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
      951  gaAAGAAGTG ACGGaGTTGA TGTGcgaAgG aCTATGTaAA AAacGGCTTG
25    1001  CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGCAGGAAG TTCAGCAAAG
      1051  CGCGCAgCAA CATTCCGACA GGGCGcCAAG TTGCCACATT GGGCGGAAAA
      1101  CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAACGCG GGAACCGGTT
      1151  TGAAGGAATC GGaCGGGGCG GTGGTCCGAT CGGCAAACTG A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

      1  MAEICLITGT PGSGKTLKMV SMANDEMFK PDEKAIRRKV FTNIKGLKIP
30     51  HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVPWPAR
      101  SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
      151  KMGMRLLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVKNV
      201  KRSKWFYTLF VIVLLIPVfV GLSYKMLSSY GKKQEEPAAQ ESAATEQQAV
      251  LPDKTEGEPV NNGNLTADMF VPTLSEKFXS KPIYNGVRQV RTEFYIAGCI
35     301  EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPKYE ESQGQEVQQS
      351  AQQHSDRAQV ATLGKPKPXQN LMYDNWEERG KPFEIGIGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

      1  ATGGCAGAGA TCTGTTTGTAT AACCGGCACG CCCGGTTCAG GGAAAACATT
40     51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
      101  ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAGGCTT GAAAATACCG
      151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
      201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
      251  TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
45     301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
      351  ACATCAGGGC ATTGATATAT TTGTTTTGAC TCAAGGTCCT AAGCTTCTAG
      401  ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACAT CGCTTCAAAC
      451  AAGATGGGTA TGCGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
      501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
50     551  AAGTTTATGA CTTGTACGAA TCACGCGGAAG TTCATACCGT AAATAAGGTC
      601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCA GTAATAGTAT TGCTGATTCC
      651  CGTGTTTGTC GGCCTGTCCT ATAAAATGTT GAGCAGTTAC GGAAAAAAAC
      701  AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
      751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
55     801  AGATATGTTT GTTCCGACAT TGTCCGAAA ACCCGAAAGC AAGCCGATT
      851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
      901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGGCATT
      951  GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTGC
60    1001  CGTTTAACCC ATACAAAGAA GAAAGCCAAG GGCAGGAAGT TCAGCAAAAGC
      1051  GCGCAGCAAC ATTCGGACAG GGCGCAAGTT GCCACATTGG GCGGAAAACC
      1101  GTAGCAGAAC CTAATGTACG ATAATTGGGA AGAACGCGGG AAACCGTTTG
      1151  AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

-215-

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK
 5 201 KRSKWFYTLF VIVLLIPVFLV GLSYKMLSSY GKKQEEPAAQ ESAATEQQAV
 251 LPDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTFEYIAGCI
 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGOEVQQS
 351 AQQHSRAQV ATLGKGP*QN LMYDNWEERG KPFEGIGGGV VGSAN*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N. meningitidis*:

15	orf84.pep	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENKIRRKV FTNIKGLKIP HTYIETDAKK
	orf84a	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK
20	orf84.pep	LPKSTDEQLSAHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
	orf84a	LPKSTDEQLSAHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
25	orf84.pep	IDIFVLTQGP KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVK MASSAFSSIYT
	orf84a	IDIFVLTQGS KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVK MASSAFSSIYT
30	orf84.pep	LDKKVYDLYE SAEVHTVNVK VRSKWFYTLF VIVLLIPVFLV GLSYKMLSSY GKKQEEPAAQ
	orf84a	LDKKVYDLYE SAEVHTVNVK VRSKWFYTLF VIVLLIPVFLV GLSYKMLSSY GKKQEEPAAQ
35	orf84.pep	ESAATEQQAV LPDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTFEYIAGCI
	orf84a	ESAATEHQAV FQDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTFEYIAGCV
40	orf84.pep	EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGOEVQQS AQQHSRAQV
	orf84a	EGGRTGCTCY SHQGTALKEIT KEMCKDYARNGL PFNPYKEESQGRDVQSQEQHSDRPQV
45	orf84.pep	ATLGKGPXQN LMYDNWEERG KPFEGIGGGV VGSANX
	orf84a	ATLGKGPWQN LMYDNWQERG KPFEGIGGGV VGSANX

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

55 1 ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
 51 AAAAATGGTT TCCATGATGG CAAACGATGA AATGTTTAAG CCGGATGAAA
 101 ACGGCATACG CCGTAAAGTA TTTACGAACA TCAAAGGCTT GAAGATACCG
 151 CACACCTACA TAGAAACGGA CGCGAAAAAG CTGCCGAAAT CGACAGATGA
 60 201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
 251 TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
 351 ACATCAGGGC ATTGATATAT TTGTTTGGAC TCAAGGCTCT AAGCTTCTAG
 401 ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACAT CGCTTCAAAC
 451 AAGATGGGTA TCGGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC

-216-

501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
 551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
 601 AAGCGGTCAA AATGGTTTTA TACTCTGCCA GTAATAATAT TGCTGATTCC
 651 CGTTTTTGTC GGCCTGTCCT ATAAAATGTT AAGTAGTTAT GGAAAAAAC
 701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA TCAGGCAGTA
 751 TTTCAGGATA AAACAGAAGG CGAGCCGGTA AACACGGTA ACCTTACCGC
 801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT
 851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA
 901 GAAGGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGGCATT
 951 GAAAGAAATT ACAAAGGAAA TGTGCAAGGA TTACGCAAGA AACGGATTGC
 1001 CGTTTAACCC ATATAAGAA GAAAGCCAAG GGCGGGATGT CCAGCAAAGT
 1051 GAGCAGCACC ATTCGGACAG ACCGCAAGTT GCCACGTTGG GCGGAAAGCC
 1101 GTGGCAAAAT CTTATGTATG ATAATTGGCA GGAGCGCGGA AAACCGTTTG
 1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

15 This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRV FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMEWIKK PENIGSIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGS KLLDQNLRTL VRKHYHIASN
 151 KMGMRLLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNKV
 201 KRSKWFTLP VIILLIPVFV GLSYKMLSSY GKKQEEPAAQ ESAATEHQAV
 251 FQDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCV
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVOQS
 351 EQHHSRDPQV ATLGKQPWQN LMYDNWQERG KPFEGIGGGV VGSAN*

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

25	orf84a.pep	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGIRRV	FTNIKGLKIP	HTYIETDAKK
	orf84-1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGIRRV	FTNIKGLKIP	HTYIETDAKK
30	orf84a.pep	LPKSTDEQLS	AHDMEWIKK	PENIGSIVIV	DEAQDVWPAR	SAGSKIPENV	QWLNTHRHQG
	orf84-1	LPKSTDEQLS	AHDMEWIKK	PENIGSIVIV	DEAQDVWPAR	SAGSKIPENV	QWLNTHRHQG
35	orf84a.pep	IDIFVLTQGS	KLLDQNLRTL	VRKHYHIASN	KMGMRLLLEW	KICADDPVKM	ASSAFSSIYT
	orf84-1	IDIFVLTQGS	KLLDQNLRTL	VRKHYHIASN	KMGMRLLLEW	KICADDPVKM	ASSAFSSIYT
40	orf84a.pep	LDKKVYDLYE	SAEVHTVNKV	KRSKWFTLP	VIILLIPVFV	GLSYKMLSSY	GKKQEEPAAQ
	orf84-1	LDKKVYDLYE	SAEVHTVNKV	KRSKWFTLP	VIILLIPVFV	GLSYKMLSSY	GKKQEEPAAQ
45	orf84a.pep	ESAATEHQAV	FQDKTEGEPV	NNGNLTADMV	VPTLSEKPES	KPIYNGVRQV	RTFEYIAGCV
	orf84-1	ESAATEHQAV	FQDKTEGEPV	NNGNLTADMV	VPTLSEKPES	KPIYNGVRQV	RTFEYIAGCV
50	orf84a.pep	EGGRTGCTCY	SHQGTALKEI	TKEMCKDYAR	NGLPFNPYKE	ESQGRDVOQS	EQHHSRDPQV
	orf84-1	EGGRTGCTCY	SHQGTALKEI	TKEMCKDYAR	NGLPFNPYKE	ESQGRDVOQS	EQHHSRDPQV
55	orf84a.pep	ATLGKQPWQN	LMYDNWQERG	KPFEGIGGGV	VGSANX		
	orf84-1	ATLGKQPWQN	LMYDNWQERG	KPFEGIGGGV	VGSANX		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N.*

gonorrhoeae:

5	orf84.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDEKAIRRKVFTNIKGLKI PHTYIETDAKK	60
	orf84ng	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGVRKVFTNIKGLKI PHTHIETDAKK	60
10	orf84.pep	LPKSTDEQLSAHDMYEWIKK PENIGSIVIVDEAQDVWPARSAGSKI PENVQWLNTHRHQG	120
	orf84ng	LPKSTDEQLSAHDMYEWIKK PENVGAIIVIVDEAQDVWPARSAGSKI PENVQWLNTHRHQG	120
15	orf84.pep	IDIFVLTQGP KLLDQNLRLTLVRKHYHIASNKMGMRLLLEWKICADDPVKMASSAFSSIYT	180
	orf84ng	IDIFVLTQGP KLLDQNLRLTLVKRHYHIAANKMGLRLTLEWKVCADDPVKMASSAFSSIYT	180
20	orf84.pep	LDKKVYDLYXXAEVHTVNKVKRSKWFTLPVIVLLIPV FVGLSYKMLSSY GKKQEEPAAQ	240
	orf84ng	LDKKVYDLYESAEIHTVNKVKRSKWFTLPV IILLIPLFVGLSYKMLGSY GKKQEEPAAQ	240
25	orf84.pep	ESAATEQQAVLPDKTEGE PVNNGNLTADMFVPTLSEKPKSKPIYNGVRQVRTFEYIAGCI	300
	orf84ng	ESAATEQQAVLPDKTEGESVNNGNLTADMFVPTLPEKPESKPIYNGVRQVRTFEYIAGCI	300
30	orf84.pep	EGGRTGCACYS HQGTALKEVTELMCKDVKNGLPFNPYKEESQGOEVQSSAQQHS DRAQV	360
	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDVKNGLPFNPYKEESQGOEVQSSAQQHS DRAQV	360
35	orf84.pep	ATLGGKQPQNLMYDNWEERGKPFEGIGGGVVG SAN 395	
	orf84ng	ATLGGKQPQNLMYDNWEERGKPFEGIGGGVVG SAN 395	

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

	1	ATGGCAGAAA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTCAG	GGAAAAACATT
35	51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAAG	CCAGATGAAA
	101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAGATACCG
	151	CACACCCACA	TAGAAACAGA	CGCAAAGAAG	CTGCCGAAAT	CAACCGATGA
	201	ACAGCTTTCG	GCGCATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg
40	251	tccgcgCAAT	CGTTATTGTC	GATGAGGCGC	AAGACGTATG	GCCCGCACGC
	301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG
	351	GCATCAGGCG	ATAGATATAT	TTGTATTGAC	ACAAGGTCCT	AAACTCTTAG
	401	ATCAGAACTT	GCGAACATTG	GTTAAAAGAC	ATTACCACAT	TGCGGCCAAC
45	451	AAAATGGGTT	TGCGTACCCT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC
	501	GGTAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA
	551	AAGTTTATGA	CTTGTACGAA	TCCGCAGAAA	TTACACGGT	AAACAAAGTC
	601	AAGCGTTCAA	AATGGTTTTA	TGCATTGCCC	GTCATCATAT	TATTGATTCC
50	651	GCTATTGTGC	GSTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAAC
	701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CAACAGAACA	GCAGGCAGTA
	751	CTTCCGGATA	AAACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC
	801	AGATATGTTT	GTTCCGACAT	TGCCCGAAAA	ACCCGAAAGC	AAGCCGATTT
55	851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
	901	GAAGGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT
	951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC
	1001	CGTTTAAACC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
60	1051	GCGCAGCAAC	ATTCCGCAGC	GGCGCAAGT	GCCACCTTGG	GCGGAAAAACC
	1101	GCAGCAGAAC	CTAATGTACG	ACAATTGGGA	AGAACGCGGG	AAACCGTTTG
	1151	AAGGAATCGG	CGGGGGCGTG	GTCGGATCGG	CAAACCTGA	

This encodes a protein having amino acid sequence <SEQ ID 326>:

	1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGVRKV	FTNIKGLKIP
60	51	HTHIETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENVGAIIV	DEAQDVWPAR
	101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGP	KLLDQNLRTL	VKRHYHIAAN
	151	KMGLRTLLEW	KVCADDPVKM	ASSAFSSIYT	LDKKVYDLYE	SAEIHTVNKV
	201	KRSKWYFALP	VIILLIPLFV	GLSYKMLGSY	GKKQEEPAAQ	ESAATEQQAV
	251	LPDKTEGESV	NNGNLTADMF	VPTLPEKPES	KPIYNGVRQV	RTFEYIAGCI
	301	EGGRTGCTCY	SHQGTALKEV	TELMCKDYVK	NGLPFNPYKE	ESQGOEVQSS
	351	AQQHSDRAQV	ATLGGKQPQN	LMYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

		10	20	30	40	50	60
	orf84-1.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRVFTNIKGLKIPHTYIETDAKK					
5	orf84ng	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGVRKVFTNIKGLKIPHTHIETDAKK					
		10	20	30	40	50	60
	orf84-1.pep	LPKSTDEQLSAHDMYEWIKKPKENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
10	orf84ng	LPKSTDEQLSAHDMYEWIKKPKENVGAIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
		70	80	90	100	110	120
	orf84-1.pep	LPKSTDEQLSAHDMYEWIKKPKENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
	orf84ng	LPKSTDEQLSAHDMYEWIKKPKENVGAIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
		70	80	90	100	110	120
	orf84-1.pep	IDIFVLTQGPVKLLDQNLRTLVRKHYHIAASNKMGMRTLLEWKICADDPVKMASSAFSSIYT					
15	orf84ng	IDIFVLTQGPVKLLDQNLRTLVRKHYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIYT					
		130	140	150	160	170	180
	orf84-1.pep	IDIFVLTQGPVKLLDQNLRTLVRKHYHIAASNKMGMRTLLEWKICADDPVKMASSAFSSIYT					
	orf84ng	IDIFVLTQGPVKLLDQNLRTLVRKHYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIYT					
		130	140	150	160	170	180
	orf84-1.pep	LDKKVYDLYESAIEVHTVNVKVRKSKWFYTLPIVIVLLIPVFLVGLSYKMLSSYGKKQEEPAQAQ					
20	orf84ng	LDKKVYDLYESAIEHTVNVKVRKSKWFYALPVIILLIPLFVGLSYKMLSGYGKKQEEPAQAQ					
		190	200	210	220	230	240
	orf84-1.pep	LDKKVYDLYESAIEVHTVNVKVRKSKWFYTLPIVIVLLIPVFLVGLSYKMLSSYGKKQEEPAQAQ					
	orf84ng	LDKKVYDLYESAIEHTVNVKVRKSKWFYALPVIILLIPLFVGLSYKMLSGYGKKQEEPAQAQ					
		190	200	210	220	230	240
25	orf84-1.pep	ESAATEQQAVLPDKTEGEPVNNNGNLTADMFPVPTLSEKPEPKPIYNGVRQVRTFEYIAGCI					
	orf84ng	ESAATEQQAVLPDKTEGESVNNNGNLTADMFPVPTLPEKPEPKPIYNGVRQVRTFEYIAGCI					
		250	260	270	280	290	300
	orf84-1.pep	ESAATEQQAVLPDKTEGEPVNNNGNLTADMFPVPTLSEKPEPKPIYNGVRQVRTFEYIAGCI					
	orf84ng	ESAATEQQAVLPDKTEGESVNNNGNLTADMFPVPTLPEKPEPKPIYNGVRQVRTFEYIAGCI					
		250	260	270	280	290	300
	orf84-1.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSQSDRAQV					
35	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSQSDRAQV					
		310	320	330	340	350	360
	orf84-1.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSQSDRAQV					
	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSQSDRAQV					
		310	320	330	340	350	360
	orf84-1.pep	ATLGGKPKXQNLMYDNWEERGKPFEGIGGGVVGSA					
40	orf84ng	ATLGGKPKQNLMYDNWEERGKPFEGIGGGVVGSA					
		370	380	390			

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 39

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

	1	GTGGTTTTTC	TGAATGCCGA	CAACGGGATA	TTGGTTCAGG	ACTTGCCTTT
50	51	TGAAGTCAAA	CTGAAAAAAT	TCCATATCGA	TTTTTACAAT	ACGGGTATGC
	101	CGCGTGATTT	CGCCAGCGAT	ATTGAAGTGA	CGGACAAGGC	AACCGGTGAG
	151	AAACTCGAGC	GCACCATCCG	CGTGAACCAT	CCTTTGACCT	TGCACGGCAT
	201	CACGATTTAT	CAGGCGAGTT	TTGCCGACGG	CGGTTCCGGAT	TTGACATTCA
	251	AGGCGTGGAA	TTTGGGTGAT	GCTTCGCGCG	AGCCTGTCGT	GTTGAAGGCA
55	301	ACATCCATAC	ACCAGTTTCC	GTTGGAAATT	GGCAAACACA	AATATCGTCT
	351	TGAGTTCGAT	CAGTTCACCT	CTATGAATGT	GGAGGACATG	AGCGAGGGCG
	401	CGGAACGGGA	AAAAAGCCTG	AAATCCACGC	TGCCCGATGT	CCGCGCCGTT
	451	ACTCAGGAAG	GTCACAAATA	CACCAAT...TACCG
	501	TATCCGTGAT	GCGCCAGGCC	AGGCGGTCGA	ATATAAAAC	TATATGCTGC
60	551	CGGTTTTGCA	GGAACAGGAT	TATTTTGGGA	TTACCGGCAC	GCGCAGCGC.

5
10

```

601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
651 AGCGGACACC TTTATGGCAT TGCCTGAGTT TTTGAAAGAT GGGGAAGGGC
701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
751 GAACAATTCA TGCTGGCTGC GGAAAACACG CTGAACATCT TTGCACAAA
801 AGGCTATTG GATTGGACG AATTTATTAC GTCCAATATC CCGAAAGAGC
851 AGCAGGATAA GATGCAGGGC TATTTCTACG AAATGCTTTA CGGCGTGATG
901 AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
951 GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTCGGG TTTGCAGATG ACCCGTTCCC C.GGTCCGCT
1101 TTTGGTCTAT CTC...
```

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

15
20

```

1 MVFLNADNGI LVQDLPEFVK LKKEHIDFYN TGMPRDFASD IEVTDKATGE
51 KLERTIRVNH PLTLHGITIY QASFADGGSD LTFKAWNLDG ASREPVVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRAV
151 TQEGHKYTNX XXXXXYRIRD APGQAVEYKN YMLPVLQEQD YFWITGTRSX
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRXVAD ATKGAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKEQQDKMQG YFYEMLYGVM
301 NAALDETXTR YGLPEWQQDE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
351 SEVRSSGLQM TRSXGPLLVY L...
```

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

25
30
35
40
45
50
55
60

```

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCCTC CGTGGTTCGC
51 TTTTTCAGC TCCATGCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT
151 TATTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTTG GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCAGTTTG TGCTGATTG GCAATGTGCC GCCGTCTCTG
301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA
401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTGCGCTG GGCGGTTGA
551 TAGACAGTAA CCTGCTGTTG AAACCTGGGT TGCTGACCGG TCGGATTGTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGTGCGCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCGGA TGTGGTTTTT CTGAATGCCG ACAACGGGAT ATTGGTTTCA
751 GACTTGCCCT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTTACAA
801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAGG
851 CAACCGGTGA GAAACCTGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTGCGGACG GCGGTTGCGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCT
1001 TGTTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTCTG AATATAAAAA
1251 CTATATGCTG CCGGTTTTTG AGGAACAGGA TTATTTTTTG ATTACCGGCA
1301 CGCGCAGCGG CTGCGAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGGAAGGG CGCAAACGTC TGTTTGCCGA CGCAACCAAA GGCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CCGAAAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTCAGAT GACCCGTTCC
1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GGCGTGGGTA TTGTTTTCAG
1901 ACGGCAAAAT CCGTTTGGCC ATGTCTTCGG CCCGAGCGA ACGGGATTTG
1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA
```

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

65

```

1 MSKSRSPPL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPTD
51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
```

-220-

5
10
15
20
25
30
35
40
45
50
55
60
65

101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVQ GFQKGTINRE
151 DGSVLIAAKK GTMNKWDYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
201 PDNQAVYAKD FKPEISILGAS NLSFRGNVNI SEGQSADVVV LNADNGILVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITYIQAS FADGSDLTG KAWNLDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLQQ QYRWLRIPLD
451 KQLKADTFMA LREFLKDGEK RKRLVADATK GAPAEIREQF MLAAENTLNI
501 FAQKGYLGLD EFITSNIPKE QODKMQGYFY EMLYGVNNA LDETIRRYGL
551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTSR
601 PGALLVYLG VLLVLGTVLM FYVREKRAWV LFSGDKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

	orf88.pep				10	20	30
					MVFLNADNGILVQDLDPFEVKLKKFHIDFYN		
20	orf88a	AKDFKPESILGASNLSFRGNVNI	SEGQSADVVFLNADNGILVQDLDPFEVKLKKFHIDFYN				
		210	220	230	240	250	260
	orf88.pep		40	50	60	70	80
25		TGMPRDFASDIEVTDKATGEKLERTIRVNHPLTLHGITYIQASFADGSDLT	FKAWNLDG				
	orf88a	TGMPRDFASDIEVTDKATGEKLERTIRVNHPLTLHGITYIQASFADGSDLT	FKAWNLDG				
		270	280	290	300	310	320
	orf88.pep		100	110	120	130	140
30		ASREPVLKATSIHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLPDVR	AV				
	orf88a	ASREPVLKATSIHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLNDVR	AV				
		330	340	350	360	370	380
35	orf88.pep		160	170	180	190	200
		TQEGHKYTNXXXXXXIRIDAPQAVEYKNYMLPVLQEQDYFWITGTRSXLLQQYRWLR	I				
	orf88a	TQEGHKYTNIGPSIVYRIRDAAGAVEYKNYMLPVLQEQDYFWITGTRSGLQQYRWLR	I				
		390	400	410	420	430	440
40	orf88.pep		220	230	240	250	260
		PLDKQLKADTFMALREFLKDGEGRKRXVADATKGAPAEIREQFMLAAENTLNI	FAQKGYL				
	orf88a	PLDKQLKADTFMALREFLKDGEGRKRLVADATKGAPAEIREQFMLAAENTLNI	FAQKGYL				
45		450	460	470	480	490	500
	orf88.pep		280	290	300	310	320
		GLDEFITSNIPKEQQDKMQGYFYEMLYGVNNAALDETXYTRYGLPEWQQDEARNR	FLLHSM				
50	orf88a	GLDEFITSNIPKEQQDKMQGYFYEMLYGVNNAALDETIRRYGLPEWQQDEARNR	FLLHSM				
		510	520	530	540	550	560
	orf88.pep		340	350	360	370	
55		DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTSRSGP	LLVYL				
	orf88a	DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTSRPGALLVYLGSVLLVLGTVLMFYVREKR					
		570	580	590	600	610	620
60	orf88a	AWVLFSDGKIRFAMSSARSERDLQKEFPKHVESLQRLGKDLNHDX					
		630	640	650	660	670	

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTCGC
51 TTTTTCAGC TCCATGCGCT TTGCGGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT

-221-

151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGG GTTTTCTGGG
 201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCTGTTATC ATGATGTTTT
 251 TGGTGGTTTC TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTCTCTGG
 301 CGCGAAATGA AGTCTTTTCG GGAAGAGGTT AAAGAAAAAT CTCTGGCGGC
 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA
 401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
 451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
 501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTGCTGCTG GCGGGTTGA
 551 TAGACAGTAA CCTGCTGTTG AACTGGGTA TGCTGACCGG TCGGATTGTT
 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
 651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
 701 AGAGTGCGGA TGTGGTTTTT CTGAATGCCG ACAACGGGAT ATTGGTTCAG
 751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTTACAA
 801 TACGGGTATG CCGCGCGATT TTGCCAGTGA TATTGAAGTA ACGGATAAGG
 851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
 901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTGGA
 951 TTTGACATT C AAGGCGTGG AATTGGGTGA TGCTTCGCGC GAGCCTGTCTG
 1001 TGTTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAT TGGCAAACAC
 1051 AAATATCGTC TTGAGTTTGA TCAGTTTACT TCTATGAATG TGGAGGACAT
 1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACATG
 1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CCGCCCTTCC
 1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTCTG AATATAAAAA
 1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGGCA
 1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGAC
 1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
 1401 TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GGCACACCTG
 1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAACAC GCTGAACATC
 1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
 1551 CCGGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
 1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
 1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
 1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTCAGAT GACCCGTTCC
 1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
 1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTTCAG
 1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCCGACGCGA ACGGGATTTG
 1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
 2001 CTTGAATCAT GACTGA

This encodes a protein having amino acid sequence <SEQ ID 332>:

40 1 MSKSRRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
 51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
 101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVQ GFQGKTINRE
 151 DGSVLIAAKK GTMNKGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
 201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
 45 251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRNVNPLT
 301 LHGITIYQAS FADGGSDLTF KAWNLDGASR EPVVLKATSI HQFPLEIGKH
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
 401 IVYRIRDAAG QAVEYKNYML PVLQEQQDYFW ITGTRSGLQQ QYRWLRIPLD
 451 KQLKADTFMA LREFLKDGEK RKRLVADATK GAPAEIREQF MLAAGENTLNI
 50 501 FAQKGYLGLED EFITSNIPKE QQDKMQGYFY EMLYGVNMNA LDETIRRYGL
 551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
 601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFS DGKIRFA MSSARSERDL
 651 QKEFPKHVES LQRLGKDLNH D*

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

55 orf88a.pep MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60
 orf88-1 MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60
 60 orf88a.pep QIFGFLGLYDVYASAWFVVI MMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
 orf88-1 QIFGFLGLYDVYASAWFVVI MMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
 orf88a.pep SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMTNKWGYIFAHVALIVICL 180
 orf88-1 SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMTNKWGYIFAHVALIVICL 180
 65 orf88a.pep GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240

	orf88-1	 GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
5	orf88a.pep	LNADNGILVQDLPEFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88-1	 LNADNGILVQDLPEFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88a.pep	LHGITIYQASFADGGSDLTFAKAWNLDASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	360
10	orf88-1	 LHGITIYQASFADGGSDLTFAKAWNLDASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
15	orf88-1	 SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88a.pep	PVLQEODYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	 PVLQEODYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
20	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88-1	 GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
25	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88-1	 LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88a.pep	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
30	orf88-1	 PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88a.pep	LQRLGKDLNHD 672	
35	orf88-1	 LQRLGKDLNHD 672	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N. gonorrhoeae*:

40	orf88.pep	MVFLNADNGILVQDLPEFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88ng	 MVFLNADNGMLVQDLPEFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88.pep	PLTLHGITIYQASFADGGSDLTFAKAWNLDASREPVVLKATSIHQFPLEIGKHKYRLEFD	120
45	orf88ng	 PLTLHGITIYQASFADGGSDLTFAKAWNLDASREPVVLKATSIHQFPLEIGKHKYRLEFD	120
	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRVAVTQEGHKYTNXXXXXXYRIRDAPGQAVEYKN	180
50	orf88ng	 QFTSMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
	orf88.pep	YMLPVLQEODYFWITGTRSLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRXVAD	240
	orf88ng	 YMLPILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVAD	240
55	orf88.pep	ATKGAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM	300
	orf88ng	 ATKDAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVM	300
60	orf88.pep	NAALDETXYTRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88ng	 NAALDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88.pep	TRSXGPLLVL	371
65	orf88ng	 TRSPGALLVYLGSVLLVLGTVFMFYVPPKRAWVLF SNXKIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

```

      1  MVFLNADNGM LVQDLPFEEVK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
      51  KLERTIRVNH PLTLHGITIY QASFADGGSD LTFKAWNLRD ASREPVLVKA
101  TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLNDVRAV
151  TQEGKKYTNi GPSIVYRIRD AAGQAVEYKN YMLPILQDKD YFWLTGTRSG
201  LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRLVAD ATKDAPAEIR
251  EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKGQODKMOG YFYEMLYGVM
301  NAALDETIIR YGLPEWQQDE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
10 351  SEVRSSGLQM TRSPGALLVY LGSVLLVLGT VFMFYVPKKR AWWLFSNXKI
401  RFAMSSARSE RDLQKEFPKH VESLQRLGKD LNHD*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

```

      1  ATGAGTAAAT CCCGTATATC TCCCACACTT CTTTCCCGTC CGTGGTTTCGC
      51  TTTTTCAGC TCCATCGCGT TTGCGGTCGC TTTGCTCAGT CTGCTGGGTA
15 101  TTGCATCGGT TATCGGCACG GTGTTACAGC AAAACCAGCC GCAGACGGAT
151  TATTTGGTCA AATTCGGACC GTTTTGACT CGGATTTTGT ATTTTTTGGG
201  TTTGTATGAT GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTC
251  TGGTGGTTTC TACCAAGTTT TGTTTAATCC GTAACGTTC CCGCTTTTGG
301  CGCGAAATGA AGTCTTTCCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
20 351  GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCCCCC GAAGTTGCCA
401  AACGTTATCT GGAGGTGCGG GGTTTTCAGG GAAAAACCGT CAGCCGTGAG
451  GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACaatga acaaATGGGG
501  CTATATCTTT GCcCaagtag ctTTGATTGT CATTTGCCTG GCGGGTTGA
25 551  TAGACAGTAA CCTGCTGCTG AAGCTGGGTA TGCTGGCCGG TCGGATTGTT
601  CCGGACAATC AGCGGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651  GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701  AAAGTGCGGA TGTGGTTTTT CTGAATGCCG ACAACGGGAT GTTGGTTCAG
751  GACTTGCTTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTTACAA
801  TACGGGTATG CCGCGCGATT TTGCCAGCGA TATTGAAGTA ACGGACAAGG
30 851  CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901  TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA
951  TTTGACATTC AAGGCGTGGA ATTTGAGGGA TGCTTCGCGC GAACCTGTCTG
1001  TGTTGAAGGC AACCTCCATA CACCAGTTTC CGTTGGAAT CCGCAAAAC
1051  AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
35 1101  GAGCGAGGGT GCGGAACGGG AAAAAAGCCT GAAATCCACT CTGAACGATG
1151  TCCGCGCCGT TACTCAGGAA GGTAaaaaat ACACCAATAT CCGCCCTTCC
1201  ATCGTGTACC GCATCCGTGA TGcggCAGGG CAGGCGGTCTG AATATAAAAA
1251  CTATATGCTG CCGATTTTGC AGGACAAAGA TTATTTTGGG CTGACCGGCA
40 1301  CGCGCAGCGG CTTCGAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351  AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAGA
1401  TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAAA GACGCACCTG
1451  CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAATATC
1501  TTTGCGCAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551  CCCGAAAGGG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
45 1601  ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651  CCCGAATGGC AGCAGGATGA AGCGCGGAAC CGTTTCCTGC TGCACAGTAT
1701  GGATGCCTAT ACGGGGCTGA CGGAATATCC CGCGCCTATG CTGCTCCAGC
1751  TTGACGGGTT TTCCGAGGTG CGTTCCTCAG GTTTGCAGAT GACCCGTTCTG
50 1801  CCGGGTGCGC TTTTGGTCTA Tctcggctcg gtattgttgg TTTTGGgtac
1851  ggtaTttatg tTTTATGTGC GCGAAAAACG GGCGTGgta tTGTTTTcag
1901  aCGGCAAAAT CCGTTTTGCT ATGtCTTcgg CCcgagcga ACGGGATTTG
1951  cAGaaggaaT TTCCAAAACA CGtcgAGAGC CTGCAACggc tcggcaaggA
2001  CttgaatCAT GACTga

```

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

```

55      1  MSKSRIPTL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPOTD
      51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
101  REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVR GFQKTVSRE
151  DGSVLIAAKK GTMNKWGYIF AQVALIVICL GGLIDSNLLL KLGLMAGRIV
201  PDNQAVYAKD FKPESILGAS NLSFRGNVNI SEGQADVVF LNADNGMLVQ
60 251  DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301  LHGITIYQAS FADGGSIDLTF KAWNLRDASR EPVVLKATSI HQFPLEIGKH
351  KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401  IVYRIRDAAG QAVEYKNYML PILQDKDYFW LTGTRSGLOQ QYRWLRPLD
451  QQLKADTFMA LREFLKDGEG RKRLVADATK DAPAEIREQF MLAAENTLNI
65 501  FAQKGYLGLD EFITSNIPKG QQDKMQGYFY EMLYGMNAA LDETIIRRYGL

```

551 PEWQQDEARN RFLHSM DAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
 601 PGALLVYLGS VLLVLGTVM FYVREKRAWV LFS DGKIRFA MSSAR SERDL
 651 QKEFPKHVES LQRLGKDLNH D*

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

```

5      orf88-1.pep  MSKSRRSPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA  60
      orf88ng-1    MSKSRRSPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGPFWT  60

10     orf88-1.pep  QIFGFLGLYDVYASAWFVIMMFLVYSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH  120
      orf88ng-1    RIFDFLGLYDVYASAWFVIMMFLVYSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH  120

15     orf88-1.pep  SSSL DVKIAPEVAKRYLEVQGFQGTINREDG SVLIAAKKGT MNK WGYIFAQVALIVICL  180
      orf88ng-1    SSSL DVKIAPEVAKRYLEV RGFQGT V SREDG SVLIAAKKGT MNK WGYIFAQVALIVICL  180

20     orf88-1.pep  GGLIDSNLLKLGMLTGRIVPDNQA VYAKDFKPESILGASNLSFRGNVNISEGQSADVVF  240
      orf88ng-1    GGLIDSNLLKLGMLTGRIVPDNQA VYAKDFKPESILGASNLSFRGNVNISEGQSADVVF  240

25     orf88-1.pep  LNADNGILVQDLPFEV KKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT  300
      orf88ng-1    LNADNGMLVQDLPFEV KKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT  300

30     orf88-1.pep  LHGITIYQASFADGGS DLT FKA WNLGDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT  360
      orf88ng-1    LHGITIYQASFADGGS DLT FKA WNLGDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT  360

35     orf88-1.pep  SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGK KYTNIGPSIVYRIRDAAGQAVEYKNYML  420
      orf88ng-1    SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGK KYTNIGPSIVYRIRDAAGQAVEYKNYML  420

40     orf88-1.pep  PVLQEQDYFWITGTRSG LQQYRWLRIPLDKQLKADTFMALREF LKDGEGRKRLVADATK  480
      orf88ng-1    PILQDKDYFWLTGTRSG LQQYRWLRIPLDKQLKADTFMALREF LKDGEGRKRLVADATK  480

45     orf88-1.pep  GAPAEIREQFM LAAENTLNI FAKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA  540
      orf88ng-1    DAPAEIREQFM LAAENTLNI FAKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVMNAA  540

50     orf88-1.pep  LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPM LLQLDGFSEVRSSGLQMTRS  600
      orf88ng-1    LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPM LLQLDGFSEVRSSGLQMTRS  600

55     orf88-1.pep  PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSAR SERDLQKEFPKHVES  660
      orf88ng-1    PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSAR SERDLQKEFPKHVES  660

60     orf88-1.pep  LQRLGKDLNH D 671
      orf88ng-1    LQRLGKDLNH D 671

```

Furthermore, ORG88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

```

55     gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
      Score = 94.4 bits (231), Expect = 2e-18
      Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)

60     Query: 16  FAFFSSMRFAVALLSLLGIASVIG-TV LQONQPQTDYLVKFGPFWTRIFDFLGLYDVYAS  74
      + F +S++ A+ ++ +LGI S++G T ++QNQ YL +FG L L DV+ S
      Sbjct: 80  YDFLASLKLAI FIMLVLGILSMLGSTYIKQNSFEWYLDQFGYDVGIWIWKLWLVNDVFHS  139

65     Query: 75  AWFVIMMFLVYSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRHSSLLDVKIAPEVAK  134
      +++++ ++ L V+ C I+ +P W++ S +E++ + A +H + VKI P+ K
      Sbjct: 140 WYYILFIVLLAVNLIFCSIKRLPRVWKQAFS-KERILKLDEHA EKHLKPITVKI-PDKDK  197

70     Query: 135 --RYLEV RGFQGT V SREDG SVLIAAKKGT MNK WGYIFAQVALIVICL GGLIDSNLLK L  192
      ++L +GF+ V E + + A+KG ++ G +AL+VI G LID
      Sbjct: 198 VLKFL LKKGFK-VFVEEEGNKLYVFAEKGRFSRLGVYITHIALLVIMAGALID-----  249

```

Query: 193 GMLAGRIVPDNQAVYAKDFKPESILGASNLSEFRGNVNISEGQSADVVFLNADNGMLVQDL 252
 +I+G RG++ ++EG + DV+ + A+ L
 Sbjct: 250 -----AIVGV-----RSLIVAEGDTNDVMLVGAE--QKPYKL 280

Query: 253 PFEVKLKKFHIDFY---NTGMPRDEFA-----SDIEVTDKATGEKLER--TIRVNHPLT 300
 PF V L F I Y N + + FA SDIE+ + G K+E T++VN P
 Sbjct: 281 PFAVHLIDFRIKTYAEENPNVDKRFAQAVSSYESDIEIIN---GGKVEAKGTVKVNEPFD 337

Query: 301 LHGITYQASFA--DGGSDLTFKAWNLRDASREP 332
 ++QA++ DG S + + + A +P
 Sbjct: 338 FGRYRLFQATYGILDGTSGMGVIVVDRKKAHEDP 371

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 337>:

```

1  ATGATGAGTA ATAmAATGGm ACAAAAAGGG TTTACATTGA TTGmGmTGAT
51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
101 ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GycGGTATCA ACAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
201 CGATAATCAG ACCATCGAGA ACAAACTGGA AATATTGTGC TCAGGCTATA
251 AGATGAATCC GAAAATTGCC AAAAAaTATA GTGTTTCGGT AAAGTTTGTC
301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
351 GACGGGTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

```

1  MMSNXMXQKG FTLLIXMIVV AILGIISVIA IPSYXSIEK GYQSOLYTEM
51 XGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKfV
101 DKEKSRAYRL VGVPKAGTGY TlSVWMNSVG DGYKCRDAAS AQAHLETlSS
151 DVGCEAFsNR KK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 339>:

```

1  ATGATGAGTA ATAAATGGA ACAAAAAGGG TTTACATTGA TTGAGATGAT
51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
201 CGATAATCAG ACCATCGAGA ACAAACTGGA AATATTTGTC TCAGGCTATA
251 AGATGAATCC GAAAATTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC
301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
351 GACGGGTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

```

1  MMSNKMEQKG FTliEMMIVV AILGIISVIA IPSYQSYIEK GYQSOLYTEM
51 VGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKfV
101 DKEKSRAYRL VGVPKAGTGY TlSVWMNSVG DGYKCRDAAS AQAHLETlSS
151 DVGCEAFsNR KK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with Pile of *N. gonorrhoeae* (accession number Z69260).

ORF89 and Pile protein show 30% aa identity in 120a overlap:

10 ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

40 This encodes a protein having amino acid sequence <SEQ ID 342>:

45 ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

BNSDOCID: <WO_9924578A2_1_>

130 140 150 160

Homology with a predicted ORF from *N.gonorrhoeae*

ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*

5	<i>gonorrhoeae:</i>		
	orf89	MMSNXMXQKGFTLIXXMIVVAILGIIISVIAIPSYXSIEKGYQSQLYTEMXGINNISKQF	60
	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF	60
10	orf89	ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY	120
	orf89ng	ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY	120
15	orf89	TLVWMNSVGDGYKCRDAASAQAHALETLSDDVGCEAFSNRKK	162
	orf89ng	TLVWMNSVGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKK	162

The complete length ORF89ng nucleotide sequence <SEQ ID 343> is:

	1	aTGATGAGCA	ATAAAATGGA	ACAAAAAGGG	TTTACATTGA	TTGAGATGAT
20	51	GATAGTTGTC	ACGATACTCG	GCATCATCAG	CGTCATTGCC	ATACCTTCTT
	101	ATCAGAGTTA	TATTGAAAAA	GGCTATCAGT	CCCAGCTTTA	TACGGAGATG
	151	GTCGGTATCA	ACAATGTTCT	CAAAACGTTT	ATTTTGAAAA	ATCCCCAGGA
	201	CGATAATGAT	ACCCTCAAGA	GCAAACTGAA	AATATTGTGC	TACGGCTATA
	251	AGATGAATCC	GAAAAttGCC	AAAAAATATA	GTGTTTCGGt	aaggtttGTC
25	301	gatGCGGAAA	AACCAAGGGC	ATACAGGTTG	GTCGGCGTTG	CGAACGCGGG
	351	GACGGGTTAT	ACTTTGTGCG	TATGGATGAA	CAGCGTGGGC	GACGGATACA
	401	AATGCCGTGA	TGCCACTTCT	GCCCAGGCCT	ATTCGACAC	CTTGCTCCGA
	451	GATAGCGGCT	GTGAAGCTTT	CTCTAATCGT	AAAAAATAG	

This encodes a protein having amino acid sequence <SEQ ID 344>:

30

```

      1  MMSNKMEQKG  FTLIEMMIVV  TILGIISVIA  IPSYQSYIEK  GYQSQLYTEM
      51  VGINNVLKQF  ILKNPQDDND  TLKSKLKIFV  SGYKMNPKIA  KKYSVSVRFV
     101  DAEKPRAYRL  VGVPNAGTGY  TLSVWMNSVG  DGYKCRDATS  AQAYSDTLSA
     151  DSGCEAFSNR  KK*

```

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng and ORF89-1 show 88.3%

35 identity in 162 aa overlap:

		10	20	30	40	50	60
	orf89-1.pep	MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSQ	LYTEMVGINNISKQF				
	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSQ	LYTEMVGINNVLKQF				
40		10	20	30	40	50	60
		70	80	90	100	110	120
	orf89-1.pep	ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKS	RAYRLVGVPKAGTG				
45	orf89ng	ILKNPQDDNDTLKSKLKFVSGYKMNPKIAKKYSVSVRFVDAEK	PRAYRLVGVPNAGTG				
		70	80	90	100	110	120
		130	140	150	160		
	orf89-1.pep	TLSVWMNSVGDGYKCRDAASAQAHALETLS	DVGCEAFSNRKKX				
50	orf89ng	TLSVWMNSVGDGYKCRDATSAQAYS	DTLSADSGCEAFSNRKKX				
		130	140	150	160		

Based on this analysis, including the gonococcal motifs and the homology with the known Pile protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that

5 ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
251 AACAAGCGTT GGCCn.AGAA TTTCAACCC...
```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRONATQVLS ILKNGDANTA
51  RQKAEAYAIP YDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
251 AACAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
501 CGTGTACCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GGCGTGGACG
551 GACTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAATA A
```

30 This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRONATQVLS ILKNGDANTA
51  RQKAEAYAIP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKLKNAN VNVKDNPVN KGGKEIIVRA EVGVPQKPV NMDFTTYQSG
151 GKYRTYNVAI EGASLTVYR NQFGEI I KAK GVDGLIAELK AKNGGK*
```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N. meningitidis*:

```

40      10      20      30      40      50      60
    orf91.pep  MKKSSLISALGIGILSIGMAFAAPADAVSQIRONATQVLSILKNGDANTARQKAEAYAIP
    orf91a     MKKSSFISALGIGILSIGMAFAAPADAVNQIRONATQVLSILKSGDANTARQKAEAYAIP
              10      20      30      40      50      60

45      70      80      90
    orf91.pep  YDFQRM TALAVGNPWXTXSDXQKQALAXEFQP
    orf91a     YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPVN
              70      80      90      100      110      120
```

orf91a KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
130 140 150 160 170 180

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

5      1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
      51 CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAACCAA ATCCGTCAAA
     101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CAACACCGCC
     151 CGCCAAAAAG CCGAAGCCTA TGCATTCCC TATTTCGATT TCCAACGTAT
     201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
     251 AACAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
     301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
     351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
     401 TACCGGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
     451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
     501 CGTGATCCGC AACCAATTGC GCGAAATTAT CAAAGCGAAA GGCGTGACG
     551 GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A

```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

20      1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
      51 RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
     101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
     151 GKRTYNVAI EGASLVTVYR NQFGEI IKAK GVDGLIAELK AKNGSK*

```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

25      10      20      30      40      50      60
orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91-1 MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
      10      20      30      40      50      60

30      70      80      90      100     110     120
orf91a.pep YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91-1 YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN
      70      80      90      100     110     120

35      130     140     150     160     170     180
orf91a.pep KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91-1 KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
      130     140     150     160     170     180

40      190
orf91a.pep GVDGLIAELKAKNGSKX
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91-1 GVDGLIAELKAKNGGKX
      190

```

Homology with a predicted ORF from *N. gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

```

50      orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf91ng VKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP 60

      orf91.pep YFDFQRM TALAVGNPWXTXSDXQKQALAXEFQP 93
55      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf91ng YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKFKNATVNVKDNPIVN 120

```

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

-230-

```

1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRONATQVLT ILKSGDAASA
51 RPKAEAYAVP YFDFQRM TAL AVGNPWR TAS DAQKQALAKE FQTL LIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51 CGGCATGGCA TTGCTCTCCC CGGCCGACGC AGTGGGACAA ATCCGCCAAA
101 ACGCCACACA GGTTTTGACC ATCCTCAAAA GCGGCGACGC GGCTTCTGCA
151 CGCCCAAAAG CCGAAGCCTA TCGGTTTCCC TATTTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTTCGGCA ACCCTTGGCG TACCGCGTCC GACGCGCAAA
251 AACAAAGCGTT GGCCAAAGAA TTCAAACCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTCAA AAACGCGACC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAGGGCGGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA
401 TCCCCGGTCA GAAGCCCGTC AATATGGACT TTACCACCTA CCAAAGCGGC
15 451 GGCAATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC
501 CGTGTACCGC AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAAATA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRONATQVLT ILKSGDAASA
20 51 RPKAEAYAVP YFDFQRM TAL AVGNPWR TAS DAQKQALAKE FQTL LIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

```

25 orf91-1.pep      10      20      30      40      50      60
      MKKSSLISALGIGILSIGMAFAAPADAVSQIRONATQVLSILKNGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91ng-1      MKKSSFISALGIGILSIGMAFASPADAVGQIRONATQVLTILKSGDAASARPKAEAYAVP
      10      20      30      40      50      60

30 orf91-1.pep      70      80      90      100     110     120
      YFDFQRM TALAVGNPWR TASDAQKQALAKEFQTL LIRTYS GTMLKLKNANVNVKDNPIVN
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91ng-1      YFDFQRM TALAVGNPWR TASDAQKQALAKEFQTL LIRTYS GTMLKFKNATVNVKDNPIVN
      70      80      90      100     110     120

35 orf91-1.pep      130     140     150     160     170     180
      KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI I KAK
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91ng-1      KGGKEIVVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTVYRNQFGEI I KAK
      130     140     150     160     170     180

40 orf91-1.pep      190
      GVDGLIAELKAKNGGKX
      |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
45 orf91ng-1      GIDGLIAELKAKNGGKX
      190

```

In addition, ORF91ng-1 shows homology to a hypothetical *E.coli* protein:

```

50 sp|P45390|YRBC ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC
      REGION PRECURSOR (F211) >gi|606130 (U18997) ORF_f211 [Escherichia coli]
      >gi|1789583 (AE000399) hypothetical 24.0 kD protein in mur2-rpoN intergenic
      region [Escherichia coli] length = 211

      Score = 70.6 bits (170), Expect = 6e-12
      Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

55 Query: 59 VPYFDFQRM TALAVGNPWR TASDAQKQALAKEFQTL LIRTYS GTMLKFKNATVNVKDNPI 118
      +PY. + AL +G +++A+ AQ++A F+ L + Y + + T + P
      Sbjct: 65 LPYVQVKYAGALVLGQYYKSATPAQREAYFAAFREYLKQAYGQALAMYHGQTYQIA--PE 122

60 Query: 119 VNKGGKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQFG 174
      G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
      Sbjct: 123 QPLGDKTIVPIRVTIIDPNRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

```

Query: 175 EIIKAKGIDGLIAELKA 191
 +++ KGIDGL A+LK+
 Sbjct: 183 TLLRTKGIDGLTAQLKS 199

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 42

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

```

10      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
     101  TGATCACGCA TACCCTCATC TCAAAATACA GTTTTGgnnn nnnnnnnnnn
     151  nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
     201  CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGGCAA
     251  AAGTCATCGT CTTCCGGCAG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTTA CGCGTCCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
     451  AAAGTATAC AAAAAACCGT AGGCGAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

```

20      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX
      51  XXXXAISKSG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 357>:

```

25      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCCAAAAC GAAACCGCTA
     101  TGACCAACGCA TACCCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
     151  CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
     201  CGACCATCAG GAAGCCGCCC GCCGAAACGG CTTAACGATG CAGCCGGCAA
     251  AAGTCATCGT CTTCCGGCAG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTTA CGCGTCCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
     451  AAAGTATAC AAAAAACCGT AGGCGAATAA
  
```

35 This corresponds to the amino acid sequence <SEQ ID 358; ORF97-1>:

```

      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTHTLT SKYSFDETVS
      51  RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N.meningitidis*:

```

45      orf97.pep      10      20      30      40      50      60
      MKHILPLIAASALCISTASAHHPASEPSTQNETAMITHTLISKYSFGXXXXXXAISKSG
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf97a          MXHILPLXXASALCISTASXHPASEPQTQNETAMTHTLTISKYSFDETVSRLETAISKSG
                        10      20      30      40      50      60
  
```

-232-

		70	80	90	100	110	120
	orf97.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK					
5	orf97a	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK					
		70	80	90	100	110	120
		130	140	150	160		
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
10	orf97a	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		

The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

	1	ATGANACACA	TACTCCCCCT	GANTGNCGCA	TCCGCACTCT	GCATTTCAAC
	51	CGCTTCGGNN	CATCCTGCCA	GCGAACCGCA	AACCCAAAAC	GAAACCGCTA
15	101	TGACCACGCA	TACCCTCACC	TCAAAATACA	GTTTTGACGA	AACCGTCAGC
	151	CGCCTTGAAA	CGGCCATAAA	AAGCAAAGGG	ATGGACATTT	TTGCCGTCAT
	201	CGACCATCAG	GAAGCCGCCC	GCCGAAACGG	CTTAACGATG	CAGCCGGCAA
	251	AAGTCATCGT	CTTCGGCAGC	CCCAAAGCCG	GTACGCCGCT	GATGGTCAAA
	301	GACCCCGCCT	TCGCCCTGCA	ACTGCCCTG	CGCGTCNTCG	TTACCGAAAC
20	351	GGACGGCAAA	GTACGCGCCG	CCTATACCGA	TACGCGCGCC	CTCATCGCCG
	401	GCAGCCGCAT	CGGTTTCGAC	GAAGTGGCAA	ACACTTTGGC	AAACGCCGAA
	451	AAACTGATAC	AAAAAACCAT	AGGCGAATAA		

This encodes a protein having amino acid sequence <SEQ ID 360>:

	1	MXHILPLXXA	SALCISTASX	HPASEPQTQN	ETAMTTHLT	SKYSFDETVS
	51	RLETAIKSKG	MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTPLMVK
25	101	DPAFALQLPL	RVXVTETDGK	VRAAYTDTRA	LIAGSRIGFD	EVANTLANAE
	151	KLIQKTIGE*				

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

		10	20	30	40	50	60
30	orf97a.pep	MXHILPLXXASALCISTASXHPASEPQTQN	ETAMTTHLT	SKYSFDETVS	RLETAIKSKG		
	orf97-1	MKHILPLIAASALCISTASAH	PASEPSTQN	ETAMTTHLT	SKYSFDETVS	RLETAIKSKG	
		10	20	30	40	50	60
35		70	80	90	100	110	120
	orf97a.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK					
	orf97-1	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK					
40		70	80	90	100	110	120
		130	140	150	160		
	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
45	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N.*

gonorrhoeae:

50	orf97.pep	MKHILPLIAASALCISTASAH	PASEPSTQN	ETAMTTHLT	SKYSFGXXXXXXA	IKSKG	60
	orf97ng	MKHILPPIAASAF	CISTASAH	PAGKPPTQN	ETAMTTHLT	SKYSFDETVS	RLETAIKSKG 60
55	orf97.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK					120
	orf97ng	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK					120
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					159
60	orf97ng	VRTAYTDTRALIVGS	RISFDEVANTLANAEKLIQKTIGEX				159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

```

      1 MKHILPPIAA SAFCISTASA HPAGKPPTQN ETAMTTHTLT SKYSFDETVS
      51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
5      101 DPAFALQLPL RVLVTETDGG VRTAYTDTRA LIVGSRISFD EVANTLANAE
      151 KLIQKTVE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 363>:

```

      1 ATGAAACACA TACTCCCcct gatcgccgca TccgcactCT GCATTTCAAC
      51 CGCTTCGGCA CACCCTGCCG GCAAACCGCC CACCCAAAAC GAAACCGCTA
     10 101 TGACCACGCA CACCCTCACC TCGAAATACA GTTTTGACGA AACCGTCAGC
      151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
      201 CGACCATCAG GAAGCGGCAC GCCGAAACGG CCTGACCATG CAGCCGGCAA
      251 AAGTCATCGT CTTCGGCACG CCCAAGGCCG GTACGCCgct GATGGTCAAA
      301 GACCCCGCCT TCGCCCTGCA ACTGCCCTG CGCGTCCTCG TTACCGAAAC
     15 351 GGACGGCAAA GTACGCACCG CCTATACCGA TACGCGCGCC CTCATCGTCG
      401 GCAGCCGCAT CAGTTTTCGAC GAAGTGGCAA ACACCTTGGC AAACGCCGAA
      451 AAAGTGATAC AAAAAACCGT AGGCGAATAA

```

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

```

      1 MKHILPLIAA SALCISTASA HPAGKPPTQN ETAMTTHTLT SKYSFDETVS
     20 51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
      101 DPAFALQLPL RVLVTETDGG VRTAYTDTRA LIVGSRISFD EVANTLANAE
      151 KLIQKTVE*

```

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

```

      25      10      20      30      40      50      60
      orf97-1.pep MKHILPLIAASALCISTASAHASEPSTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG
      orf97ng-1   MKHILPLIAASALCISTASAHAGPAGKPPTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG
      30      70      80      90      100     110     120
      orf97-1.pep MDIFAVIDHQEAARRNGLTMQPAKVIVFGT PKAGTPLMVKDPAFALQLPLRVLVTETDGG
      orf97ng-1   MDIFAVIDHQEAARRNGLTMQPAKVIVFGT PKAGTPLMVKDPAFALQLPLRVLVTETDGG
      35      70      80      90      100     110     120
      orf97-1.pep VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
      orf97ng-1   VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX
      40      130     140     150     160

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described
 45 above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1.

Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

```

5      1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC
      51  GCTGATGCTC CCCGCTTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA
     101  GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
     151  CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGg
     201  CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTCCGCC CCGATAATCG
    10  251  CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACaATATT
     301  GACTACAAAC TGAGTTTCCA TCCGCTGACC AaACGCTACC GCGTTACCgT
     351  CGgCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
     401  CCGGCGCGGT TGCCAACCTG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
     451  GCGGAAGCAG GGGAAACCAA GCGGAAATC CGCCTGACGC TGTCCACTTC
    15  501  AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
     551  ATTTGGATTC GGGTTGAAA CCTCTAAACA TCATCGGGAA CAAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

```

      1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
     51  RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYRFLK QQLIGDDDDNI
    101  DYKLSFHPLT KRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
    151  AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWLHDSGWK PLNIIGNK*
  
```

Further work revealed the following DNA sequence <SEQ ID 367>:

```

      1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC
     51  GCTGATGCTC CCCGCTTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA
    101  GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
    151  CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGG
    201  CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTCCGCC CCGATAATCG
    251  CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACAATATT
    301  GACTACAAAC TGAGTTTCCA TCCGCTGACC AACCCTACC GCGTTACCGT
    351  CGGCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
    401  CCGGCGCGGT TGCCAACCTG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
    451  GCGGAAGCAG GGGAAACCAA GCGGAAATC CGCCTGACGC TGTCCACTTC
    501  AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
    551  ATTTGGATTC GGGTTGAAA CCTCTAAACA TCATCGGGAA CAAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

```

      1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
     51  RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYRFLK QQLIGDDDDNI
    101  DYKLSFHPLT NRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
    151  AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWLHDSGWK PLNIIGNK*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N.meningitidis*:

```

45      orf106.pep      10      20      30      40      50      59
      MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ
      ||||| |||:: ||::: ||||| ||||| ||||| ||||| ||||| |||||
      orf106a          10      20      30      40      50      60
      MAFITRLFKSIKQWLVLPLMLSVLPDAAEGIDVSRAEARIXDGGQLSXXSRFQTELPDQ

50      orf106.pep      60      70      80      90      100     110     119
      LQQALRRGVPLNFTLSWQLSAPIIASYRFLKGLQIGDDDDNIDYKLSFHPLTKRYRVTGVA
      || | ||| || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```


The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

30 This encodes a protein having amino acid sequence <SEQ ID 370>:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N. gonorrhoeae*:

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is

55 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

```

1   ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
51  GTTGCCGATA CTCTCCGTTT TGCCGGACGC GCGGGCGGAG GGCATTGCCG
101 CGACCCGCGC CGAAGCGAGG ATAACCGACG GCGGGCGGCT TTCCATCAGC
5   151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAACAGG CGTTGCGCCG
201 GGGCGTACCG CTCAACTTTA CCTTAAGCTG GCAGCTTTCC GCGCCGACAA
251 TCGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAAT
301 ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCCTG ACCGCGTTAC
10  351 CGTCGGCGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATTGCGCG
401 CGACCGGCGC GGTGCGCAAC TGGAAAGTCC TGAACAAAGG CGCGTTGTCC
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCCTGA CGCTGTCCAC
501 TTCAAACCTG CCCAAGCCTT TCCAAATCAA CGCATTGACT TCTCAAACCT
551 GGCATTGGA TTCGGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 372>:

```

15  1   MAFITRLFKS IKQWLVLPLI LSVLPDAAAE GIAATRAEAR ITDGGRLSIS
51  SRFQTELPDQ LQQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
101 IDYKLSFHPL TNRYRVTVGA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLDSGW KPLNIIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

1   ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
51  GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CCGCGGCGGG GCTgACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
35  201 CACCGCCGAC AAAGACAcCT TGTTCAAAAC CCTGTTCTTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCAC TCGAGATGCC gCCGCGGCa TCGGGCTGGT
351 GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTCG
401 GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGccCAAG
40  451 CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCGGTCTGA CCGCGTTTA CCGCGTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTCGCCCCG CGTCTGCAC CCGGGG.TGC GCTACGGCAT
45  651 ACCGATCGCA CTGAGCAGCA TCGCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTTCTT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTTCG
751 ATGGGTATTT CGTTCGGCGG GCGGCATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTCCGCGC AATCGAAGAA AACGCCCCCG
851 CCGCTCGCCT CTCGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC
901 GCCCTCTGC.TGACCGGCAT TTTCTCGCCC CTTGCCTCCC TCCTGCTGCC
50  951 GGAAACTAC GCGCCGCTCC GGTATTATCGT CGTATCGTGT ATG.TGCCGC

```

-237-

5
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTT
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCGTGTGCTG CTGGGGCTTG ACCGTGCCGT ACCGCGCAGG CCGCC.GGCG
1151 CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTTGC CTTCAAGACC
1201 GAAAGCTCyt GCCGCTGTG GCAGCCGCTC AAACGCCTGC CGCTTTATCT
1251 GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
1301 CGCCGGCAAA CTATCCCCTG TTGCGCGCGC TATGGGCGGC ATATCTGGCA
1351 GGCTGCATCC TGCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT
1401 GAAAAACAA GTTTCCCAT TATGA

10 This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

15
20
25
30
35
40
45
1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLEF LSFLPIRFL LVLRMERAL AFSSAQLVPK
151 LAILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGXYGIPIA LSSIAWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCXTGIFSP LASLLLPENY AAVRFIVVSC MXPLPCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLDRAVPA PXGAACVACA SFWLFFAFKT
401 ESSCRLWQPL KRLPLYLHTL FCLTSSAAYT CFGTPANYPL FAGVWAAYLA
451 GCILRHRKDL HKLFHYLKKQ GFPL*

Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

25
30
35
40
45
50
55
60
1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCTCCGCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGCGCGG GCTGACGGTG
151 TCGGTGTGTG GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAAGACACCT TGTTCAAAAC CCTGTCCTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC GCCGCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTGC
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGTCCTGA CCGCGGTTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCTT TTTGCTGTTT CAAAACCGAT GCCGCTGAA GGCCGTCGCG
601 CACGCACCGT TTTCCGCCGC CGTCTGCAC CCGGGGCTGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCTT GAAAAATAT GCCGGCTGG AACAGCTCG CGTTTTATTG
751 ATGGGTATTT CGTTCGGCGG GCGGCGATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCCG
851 CCGCCCGCCT CTCGGCAACG GCAGAAATCC CCGCCGCCCT GCTTGCCCTC
901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTTGCTTCCC TCCTGTGCC
951 GGAAACTAC GCCGCCGTCC GGTATTATCGT CGTATCGTGT ATGCTGCCGC
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTGCCGTC CGGCGGCGCG CGCGCGCGG
1151 CGGTTGCCCTG TGCCGCCCTCA TTCTGGCTGT TTTTGCCTT CAAGACCGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCG TTTATCTGCA
1251 CACATTGTTT TGCCTGACCT CCTCGGCGGC CTACACCTGC TTCGGCACGC
1301 CGGCAAACTA TCCCTGTTT GCCGGCGTAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCCTGC GCCACCGGAA AGATTTCAC AAAGTGTTC ATTATTGAA
1401 AAAACAAGGT TTCCCATAT GA

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

55
60
1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLEF LSFLPIRFL LVLRMERAL AFSSAQLVPK
151 LAILLLLPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGLRYGIPIA LSSIAWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPLPCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAACVACA SFWLFFAFKTE
401 SSCRLWQPLK RPLPLYLHTL CLTSSAAYTC FGTPANYPLF AGVWAAYLAG
451 CILRHRKDLH KLFHYLKKQG FPL*

Computer analysis of this amino acid sequence gave the following results:

Prediction

ORF10-1 is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

- 5 ORF10 shows homology with the epsM gene of *S. thermophilus*, which encodes a protein of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

10 Query: 213 LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVYSMGISFGGAALLQSIFSTVW 270
L Y +PL SS+ +W L ++ R F+ + G G+ ++ + +IF+ W
Sbjct: 210 LYYALPLIPSSILWLLNASSRYFVLFLLGAGANGLLAVATKIPSIISIFNTIFTQAW 267

15 Identities = 15/57 (26%), Positives = 31/57 (54%)

Query: 7 LGYAAGSIGSAVLAVIILPLLSSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63
L + G++GS +L +++PL ++ + G L QT A L + ++ + + A +R
20 Sbjct: 12 LVFTIGNLGSKLLVFLVPLYTYAMTPQEYGMADLYQTTANLLLPLITMNVFDTLR 68

Identities = 16/96 (16%), Positives = 36/96 (37%)

Query: 307 IFSPLASLLLPEYAAVRFTVVSCLPPLFYTLTEISGIGLNVVRKTRPIXXXXXXXXXX 366
+ F+ ++ +YA+ V ML LF + ++ G ++T+ +
25 Sbjct: 305 VLKPIVEKVVSSDYASSWQYVPFMLSMLFSSFSDFGFTNYIAAKQTKGVFMTSIYGTIV 364

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF10 shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) from strain A of *N.*

meningitidis:

30		10	20	30	40	50	60
	orf10.pep	MDTKEILXYAAGSIGSAVLAVIILPLLSSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
	orf10a	MDTKEILGYAAGSIGSAVLAVIILPLLSSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
35		10	20	30	40	50	60
	orf10.pep	YVREYYATADKDTLTKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVFFE					
	orf10a	YVREYYAAADKDTLTKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVFFE					
40		70	80	90	100	110	120
	orf10.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAIIILLXPLTVGLLHFPANTAVLTAVYALA					
	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAIIILLXPLTVGLLHFPANTAVLTAVYALA					
45		130	140	150	160	170	180
	orf10.pep	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAIWGLASADRLFLKKY					
	orf10a	NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAIWGLASADRLFLKKY					
50		190	200	210	220	230	240
	orf10.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
55		250	260	270	280	290	300
	orf10.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					

-239-

```

      310      320      330      340      350      360
orf10.pep  ALCXTGIFSPLASLLLLPENYA AAVRFIVVSCMXPLFCTLA EISGIGLNVVRKTRPIALAT
5  orf10a   ALCLTGIFSPLASLLLLPENYA AAVRFIVVSCMLPPLFCTLVEISGIGLNVVRKTRPIALAT
      310      320      330      340      350      360

      370      380      390      400      410      419
orf10.pep  LGALAAANLLLLGLDRAV PAR-PXGA AVACAASF LFFAFKTESSCRLWQPLKRLPLYLHT
10 orf10a   LGALAAANLLLLGL--AVPSGGARGAAVACAASF LFFVFKTESSCRLWQPLKRLPLYMHT
      370      380      390      400      410

      420      430      440      450      460      470
orf10.pep  LFCLTSSAAYTCFGTPANYPLFAGVWAA YLAGCILRHRKDLHKL FHYLKKQGFPLX
15 orf10a   LFCLASSAAYTCFGTPANYPLFAGVWAVYLAGCILRHRKDLHKL FHYLKKQGFPLX
      420      430      440      450      460      470

```

The complete length ORF10a nucleotide sequence <SEQ ID 377> is:

```

20 1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCTGCCG
101 ACGACATCGG ACGCATCGTG CTGATGCAGA CGGCGGCGGG GCTGACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CGCCGCCGAC AAAGACACTT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC ATCCCTGCCG
25 301 TCTGAAATCC TGTTTTGCTC CGACGATGCC GCGCCCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGTCCAAG
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 GGCGAACACC GCCGTCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
30 551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCCG
601 CGCGCACCGT TTTTCATCCGC CGTCTGTCAT CGCGGCCTGC GCTACGGCAT
651 ACCGATCGCA CTAAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTTCT GAAAAATAT GCCGCGCTAG AACAGCTCGG CGTTTATTCG
35 751 ATGGGTATT CGTTCGGCGG AGCGGCATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGCA AACGCCCGCG
851 CCGCCCGCCT CTCGGCAACG GCAGAAATCCG CCGCCGCCCT GCTTGCCTCC
901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTCGCCTCCC TCCTGTGCCG
951 GGAAACTAC GCCCGCGTCC GGTTTATCGT CGTATCGTGT ATGCTGCCTC
40 1001 CGCTGTTTTG CACGCTGGTA GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGAAAAACAC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGCGTTG CCGTACCGTC CGGCGGCGCG CGCGGCGCGG
1151 CGGTTGCCTG TGCCGCTCA TTTTGGCTGT TTTTGTGTTT CAAGACCGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCGC TTTATATGCA
1251 CACATTGTTC TGCCTGGCCT CCTCGGCGGC CTACACCTGC TTCGGCACTC
45 1301 CGGCAACTA CCCCTGTTT GCCGCGGTAT GGGCGGTATA TCTGGCAGGC
1351 TGCATCCTGC GCCACCGGAA AGATTGTCAC AAAGTGTTC ATTATTTGAA
1401 AAAACAAGGT TTCCATTAT GA

```

This encodes a protein having amino acid sequence <SEQ ID 378>:

```

50 1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYAAAD KDTLFKTLFL PPLLSAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSFPIRFL LVLRMGRAL AFSSAQLVSK
151 LAIIIIIIPLT VGLLHFPANT AVLTAVYALA NLAAAALLF QNRCRLKAVR
201 RAPFSSAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSIISTVW TPYIFRAIEA NAPPARLSAT AESAAALLAS
55 301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPLFCTLV EISGIGLNVV
351 RKTRPIALAT LGALAAANLLL LGLAVPSGGA RGA AVACAAS FWLFFVFKTE
401 SSCRLWQPLK RPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAVYLAG
451 CILHRKDLH KLFHYLKKQG FPL*

```

ORF10a and ORF10-1 show 95.4% identity in 475 aa overlap:

```

60 orf10-1.pep MDTKEILXYAAGSIGSAVLA VIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
      10      20      30      40      50      60
orf10a       MDTKEILGYAAGSIGSAVLA VIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
      10      20      30      40      50      60
65

```

-240-

		70	80	90	100	110	120
	orf10-1.pep	YVREYYATADKDTL	FKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE				
5	orf10a	YVREYYAAADKDTL	FKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE				
		70	80	90	100	110	120
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	ILLXPLTVGLLHFPANTAVLTAVYALA				
10	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLA	ILLXPLTVGLLHFPANTAVLTAVYALA				
		130	140	150	160	170	180
	orf10-1.pep	NLAAAFLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSS	IAYWGLASADRLFLKKY				
15	orf10a	NLAAAFLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSS	IAYWGLASADRLFLKKY				
		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
20	orf10a	AGLEQLGVYSMGISFGGAALLFQSFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
		250	260	270	280	290	300
	orf10-1.pep	ALCXTGIFSPPLASLLLPENYA	AVRFIVVSCMXPPLFCTLA	ISGIGLN	VVRKTRPIALAT		
25	orf10a	ALCLTGIFSPPLASLLLPENYA	AVRFIVVSCMLPPLFCTLVE	ISGIGLN	VVRKTRPIALAT		
		310	320	330	340	350	360
	orf10-1.pep	LGALAANLLLLGLDRAV	PAR-PXGA	AVACAASF	WLF	FAFKTESSCRLWQPLKRLPLYLHT	
30	orf10a	LGALAANLLLLGL--AVPSGG	ARGA	AVACAASF	WLF	FVFKTESSCRLWQPLKRLPLYMHT	
		370	380	390	400	410	419
	orf10-1.pep	LFCLTSSAAYTCFGTP	ANYPLFAGVWAAYLAGCILRHRKDLHKL	FHYLKKQGFPLX			
35	orf10a	LFCLASSAAYTCFGTP	ANYPLFAGVWAVYLAGCILRHRKDLHKL	FHYLKKQGFPLX			
		420	430	440	450	460	470
	orf10-1.pep	MDTKEILGYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA				60
40	orf10a	MDTKEILXYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA				60

Homology with a predicted ORF from *N.gonorrhoeae*ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*45 *gonorrhoeae*:

	orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
50	orf10ng.pep	YVREYYAAADKDTLTKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	120
	orf10nm	YVREYYATADKDTLTKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	120
55	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTSVLTAUYALA	180
	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAUYALA	180
	orf10ng.pep	NLAAAFLFQNRCLKAVRRAPFSPAVLHRGLRYGIPLALSSLAYWGLASADRLFLKKY	240
60	orf10nm	NLAAAFLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY	240
	orf10ng.pep	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS	300
	orf10nm	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS	300
65	orf10ng.pep	ALCLTGIFSPLASLLLPENYA AVRFTVVSCLMPLFYTLTEISGIGLNVVRKTRPIALAT	360
	orf10nm	ALCXTGIFSPLASLLLPENYA AVR FIVVSCMXPPLFCTLA EISGIGLNVVRKTRPIALAT	360

-241-

```

              370      380      390      400      410
orf10ng.pep  LGALAAANLLLLGL--AVPSGGTRGA AVACAASFWLFFVKTESSCRLWQPLKRLPLYMHT
5  orf10nm    LGALAAANLLLLGLDRAV PAR-PXGA AVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT
              370      380      390      400      410

              420      430      440      450      460      470
orf10ng.pep  LFCLASSAAYTCFGTPANYPLFAGVWAA YLAGCILRHRKNLHKL FHYLKKQGFP L X
10 orf10nm    LFCLTSSAAYTCFGTPANYPLFAGVWAA YLAGCILRHRKDLHKL FHYLKKQGFP L X
              420      430      440      450      460      470

```

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

```

1  ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
15 51  GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCcccgCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG ACTGACGGTG
151 TCGGTATTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CGCGCGCGAC AAAGACACTT TGTTCAAAC CCTGTTCCTG CCGCGCTGC
251 TGTTTTCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCGCT CGACGATGCC GCCGCGGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG
401 GTATGGAAGG GCGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAA
451 CTCGCCATTC TGCTGCTGTT GCCGCTGACG GTCGGGCTGC TGCATTTCCT
501 GGCGAACACC TCCGTCCTGA CCGCCGT TTA CGCGCTGGCA AACCTTGCCG
25 551 CCGCCGCCCT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CGCGCGCCGT TTTGCGCCGC CGTCCTGCAC CGGGGGCTGC GCTACGCGAT
651 ACCGCTCGCA CTGAGCAGCC TTGCCTATTG GGGGCTGGCA TCCGCGGACC
701 GTTTGTTTCTT GAAAAAATAT GCGGGCCTGG AACAGCTCGG CGTTTATTCG
751 ATGGGTATTT CGTTCGCGCG GCGGCATTA TTGCTCCAAA GCATCTTTTC
30 801 AACGGTCTGG ACACCGTATA TTTTCCGTGC AATCGAAGAA AACGCCACGC
851 CCGCCCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTCC
901 GCCCTCTGCC TGACCGGAAT TTTCTCGCCC CTCGCCCTCC TCCTGCTGCC
951 GGAAACTAC GCGCCGTCC GGTTTACCGT CGTATCGTGT ATGCTGccgc
1001 cgcTGT TTTA CACGCTGACC GAAATCAGCG GCATCGGTTT GAACGTCGTC
35 1051 CGCAAACGC GTCCGATCGC GCTTGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTACCGTC CGGCGGCACG CGCGGCGCGG
1151 CGGTTGCCTG TGCCGCCTCA TTCTGGTTGT TTTTGT TTT CAAGACAGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCGC TTTATATGCA
1251 CACATTGTTC TGCTGCTGCT CCTCGGCGGC CTACACCTGC TTCGGCACAC
40 1301 CGGCAACTA CCCcctgtt gcccggcgtAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCCTGC GCCACCGAA AAATTGCAC AAAGTGTTC ATTATTGAA
1401 AAAACAAGGT TTCCATTAT GA

```

This encodes a protein having amino acid sequence <SEQ ID 380>:

```

1  MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
45 51  SVLCLGLDQA YVREYYAAD KDTLFKTLFL PPLLFSAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSFLPIRFL LVLMEGRAL AFSSAQLVPK
151 LAIIIIIIPLT VGLLHFPANT SVLTAVYALA NLAAAFLLF QNRCRLKA VR
201 RAPFSPAVLH RGLRYGIPLA LSSLAYWGLA SADRLFLKKY AGLEQLGVYS
50 251 MGISFGGAAL LQSIFSTVW TPYIFRAIEE NATPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFTVWSC MLPPLFYTLT EISGIGLNVV
351 RKTRPIALAT LGALAAANLLL LGLAVPSGGT RGA AVACAAS FWLFFVKTE
401 SSCRLWQPLK RLPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAA YLAG
451 CILRHRKNLH KLFHYLKKQG FPL*

```

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

```

55 10 20 30 40 50 60
orf10-1.pep MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
orf10ng-1 MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
60 10 20 30 40 50 60
70 80 90 100 110 120
orf10-1.pep YVREYYATADKDTLFKTLFLPPLLSAAIAALLLSRPSLPSEILFSLDDAAAGIGLVLE
orf10ng-1 YVREYYAADKDTLFKTLFLPPLLFSAIAALLLSRPSLPSEILFSLDDAAAGIGLVLE
65 70 80 90 100 110 120

```

-242-

		130	140	150	160	170	180
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTAVLTAVYALA					
5	orf10ng-1	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTSVLTAVYALA					
		130	140	150	160	170	180
	orf10-1.pep	NLAAAAFLLFQNRCLKAVRHAPFSPAVLHRGLRYGIPIALSSIAYWGLASADRLFLKKY					
10	orf10ng-1	NLAAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPIALSSLAYWGLASADRLFLKKY					
		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
15	orf10ng-1	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS					
		250	260	270	280	290	300
	orf10-1.pep	ALCLTGIFSPPLASLLLPENYAARFVIVVSCMLPPLFCTLAIEISGIGLNVVRKTRPIALAT					
20	orf10ng-1	ALCLTGIFSPPLASLLLPENYAARFVIVVSCMLPPLFYTLTEISGIGLNVVRKTRPIALAT					
		310	320	330	340	350	360
	orf10-1.pep	LGALAANLLLLGLAVPSGGARGAACAASFWLFFAFKTESSCRLWQPLKRLPLYLHTLF					
25	orf10ng-1	LGALAANLLLLGLAVPSGGTRGAACAASFWLFFVFKTESSCRLWQPLKRLPLYMHTLF					
		370	380	390	400	410	420
	orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLHLYLKKQGFPLX					
30	orf10ng-1	CLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLHLYLKKQGFPLX					
		430	440	450	460	470	
	orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLHLYLKKQGFPLX					
35	orf10ng-1	CLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLHLYLKKQGFPLX					
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several
 40 transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6
 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

45	1..ATCCTGAAAC	CGCATAACCA	GCTTAAGGAA	GACATCCAAC	CTGATCCGGC
	51	CGATCAAAAC	GCCTTGTC	AACCGGATGC	TGCGACAGAG
	101	CGGATGCGGA	AAATGCTGCC	GACAAGCAGC	CCGTGCGCGA
	151	GAGGTTGAAG	AAAAGGCGGG	CGAGCCGGAA	CGGGAAGAGC
	201	GGCAGTGCGT	AAGAAAGCGC	TGACGGAAGA	GCGTGAACAA
50	251	AAAAAGCGCA	GAAGAAAGAT	GCCGAAACGG	TTAAATACA
	301	CCGTCTAAAG	AAACAGAGAA	AAAAGCTTCA	AAAGAAGAGA
	351	GAAGGAAAAA	GTTGCACCCA	AACCAACCCC	GGAACAAATC
	401	GCAgCATCGA	AAAmGCGCGC	AgTGCCGCCG	CCAAAGAAGT
	451	AA.AACGTCC	GACAAGGCGG	AAGC.AACGC	ATTATCTGCA
55	501	TATGCCGACC	GTCAGAGCGC	GGAAGGGCAG	CGTGCCAAAC
	551	GGGCATATCT	TCCAAGGTGG	TCGGTTATCA	GGCGGGACAT
	601	ACCGGGTGCA	AAGCGGCAAT	ATGTCTGCCG	ATGCGGTGA

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

60	1..ILKPHNQLKE	DIQDPDADQN	ALSEPDAATE	AEQSDAENAA	DKQPVADKAD
	51	EVEEKAGEPE	REEPDGQAVR	KKALTEEREQ	TVREKAQKKD
					AETVKIQAVK

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF65 shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) from strain A of *N. meningitidis*:

```

35      orf65.pep      10      20      30
                        ILKPHNQLKEDIQPDPAQNALSEPDAATE
                        |||:| | | | | | | | | | | | | | | |
      orf65a      IIAGILFYLNQSGQNAFKIPVPSKQPAETEILKPKNQPKEDIQPEPADQNALSEPDDAAKE
                    30      40      50      60      70      80

40      orf65.pep      40      50      60      70      80      90
                        AEQSDAENAADKQPVADKADEVEEKAGEPEREEDPGQAVRKKALTEEREQTVREKAQKKD
                        |||:| | | | | | | | | | | | | | | |
      orf65a      AEQSDAEKAADKQPVADKADEVEEKAGEPEREKSDGQAVRKKALTEEREQTVGEKAQKKD
                    90      100      110      120      130      140

45      orf65.pep      100      110      120      130      140      150
                        AETVKIQAVKPSKETEKKASKEEKAAAEKVAPKPTPEQILNSGSIEXARSAAAKEVQKM
                        ||| | | | | | | | | | | | | | | | | | | | | |
      orf65a      AETVKKQAVKPSKETEKKASKEEKAAEKEKVAPKPTPEQILNSGSI EKARSAAAKEVQKM
                    150      160      170      180      190      200

50      orf65.pep      160      170      180      190      200      210
                        XNVRQGSXRIICKWARMPTVRARKGSVPNWQSWAYLPRWSVIRRDIKRFTGCKAAICLP

55      orf65a      KTPDKAEATHYLQMGAYADRRSAEGQRAKLAILGISSKVVG YQAGHKTL YRVQSGNMSAD
                    210      220      230      240      250      260

```

The complete length ORF65a nucleotide sequence <SEQ ID 385> is:

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTTTT
51 CTTTCGGTTTG ATACTGGCGA CGGTCATTAT TGCCGGTATT TTGTTTTATC

-244-

5
10
15

```

101 TGAACCAGAG CGGTCAAAAT GCGTTCAAAA TCCCGGTTCC GTCGAAGCAG
151 CCTGCAGAAA CGGAAATCCT GAAACCGAAA AACCAGCCTA AGGAAGACAT
201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGCTGCGA
251 AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT
301 GCCGACAAAG CCGACGAGGT TGAGGAAAAG GCGGACGAGC CGGAGCGGGA
351 AAAGTCGGAC GGACAGGCAG TGCGCAAGAA AGCACTGACG GAAGAGCGTG
401 AACAAACCGT CGGGGAAAAA GCGCAGAAGA AAGATGCCGA AACGGTTAAA
451 AAACAAGCGG TAAAACCATC TAAAGAAACA GAGAAAAAAG CTTCAAAAGA
501 AGAGAAAAAG GCGGAGAAGG AAAAAAGTTGC ACCCAAACCG ACCCCGGAAC
551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCTGCCAAA
601 GAAGTGCAGA AAATGAAAAC GCCCGACAAG GCGGAAGCAA CGCATTATCT
651 GCAAATGGGC GCGTATGCCG ACCGCCGGAG CGCGGAAGGG CAGCGTGCCA
701 AACTGGCAAT CTTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGGCGGGA
751 CATAAACGCG TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT
801 GAAAAAATG CAGGACGAGT TGAAAAACA TGAAGTCGCC AGCCTGATCC
851 GTTCTATCGA AAGCAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 386>:

20

```

1 MFMNKFSQSG KGLSGFFFFL ILATVIIAGI LFYLNQSGQN AFKIPVPSKQ
51 PAETELKPK NQPKEDIQPE PADQNALSEP DAAKEAEQSD AEKAADKQPV
101 ADKADEVEEK ADEPEREKSD GQAVRKKALT EEREQTVGEK AQKKDAETVK
151 KQAVKPSKET EKKASKEEKK AEKEKVAPKP TPEQILNSGS IEKARSAAAK
201 EVQKMKT PDK AEATHYLQMG AYADRRSAEG QRAKLAILGI SSKVVGQYQAG
251 HKTLYRVQSG NMSADAVKKM QDELKKHEVA SLIRSIESK*

```

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

25
30
35
40
45
50

```

      10      20      30      40      50      60
orf65a.pep MFMNKFSQSGKGLSGFFFFLILATVIIAGILFYLNQSGQNAFKIPVPSKQPAETELKPK
           |||||:|||||
orf65-1    MFMNKFSQSGKGLSGFFFFLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETELKPK
      10      20      30      40      50      60

      70      80      90     100     110     120
orf65a.pep NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADEPEREKSD
           |||||:|||||
orf65-1    NQPKEDIQPEPADQNALSEPDAAEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP
      70      80      90     100     110     120

      130     140     150     160     170     180
orf65a.pep GQAVRKKALTEEREQTVGEKAQKKDAETVKKQAVKPSKETEKKASKEEKKAEKEKVAPKP
           |||||:|||||
orf65-1    GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAAKEKVAPKP
      130     140     150     160     170     180

      190     200     210     220     230     240
orf65a.pep TPEQILNSGSIEKARSAAAKEVQKMKT PDKAEATHYLQMGAYADRRSAEQRAKLAILGI
           |||||:|||||
orf65-1    TPEQILNSGSIEKARSAAAKEVQKMKTSDKAEATHYLQMGAYADRSAEQRAKLAILGI
      190     200     210     220     230     240

      250     260     270     280     290
orf65a.pep SSKVVGQYQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX
           |||||:|||||
orf65-1    SSKVVGQYQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX
      250     260     270     280     290

```

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

60

```

      30      40      50      60      70      80
ORF65ng IIAGILLYLNQGGQNAFKIPAPSKQPAETELKLNQPKEDIQPEPADQNALSEPDVAKE
           |||:|||||:|||||:|||||:|||||:|||||
ORF65    ILKPHNQLKEDIQPD PADQNALSEPDAAE
           10      20      30

```

-245-

```

      90      100      110      120      130      140
ORF65ng AEQSDAEKAADKQPVADKADEVEEKAGEPEREEPDGQAVRKKALTEEREQTVREKAQKKD
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
ORF65   AEQSDAENAADKQPVADKADEVEEKAGEPEREEPDGQAVRKKALTEEREQTVREKAQKKD
5       40      50      60      70      80      90

      150      160      170      180      190      200
ORF65ng AETVKKKAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
ORF65   AETVKIQAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSGSIEKARSAAAKEVQKM
10      100     110     120     130     140     150

      210      220      230      240      250      260
ORF65ng KNFGQGGSQRIICKWARMPNPGARKGSVPNWQSWAYLPKWSAIRRDIKRFTACKAAICPP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
ORF65   XNVRQGGSXRIICKWARMPTVRARKGSVPNWQSWAYLPWVSVIRRDIKRFTGCKAAICLP
15      160     170     180     190     200     210

ORF65ng MR
20      ||
ORF65   MR

```

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino acid sequence <SEQ ID 388>:

```

25      1  MFMNKFSQSG KGLSGFFFGL ILATVIIAGI LLYLNQGGQN AFKIPAPSKQ
      51  PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV
      101  ADKADEVEEK AGEPEREEDP QAVRKKALT EEREQTVREK AQKKDAETVK
      151  KKAVKPSKET EKKASKEEK AAKEKVAPKP TPEQILNSRS IEKARSAAAK
      201  EVQKMKNFGQ GGSQRIICKW ARMPNPGARK GSVPNWQSWA YLPKWSAIRR
      251  DIKRFTACKA AICPPMR*

```

30 After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:

```

      1  ATGTTTATGA ACAATTTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTCTT
      51  CTTCGGTTTG ATACTGGCAA CGGTCATTAT TGCCGGTATT TTGCTTTTATC
      101  TGAAC CAGGG CGGTCAAAT GCGTTCAAAA TCCCGGCTCC GTCGAAGCAG
      151  CCTGCAGAAA CGGAAATCCT GAAACTGAAA AACCAGCCTA AGGAAGACAT
35      201  CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGTTGCGA
      251  AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT
      301  GCCGACAAag ccgacgAGGT TGAAGAAAag GcGGgcgAgc cggAACGGga
      351  aGAGCCGGAC ggACAGGCAG TGCGCAAGAA AGCACTGAcg gAAGAgcGTG
      401  AACAAACcgt cagggAAAAA GCGCagaaga AAGATGCCGA AACGgTTAAa
40      451  AAacaaGCgg tAaaaccgtc tAAAGAAACa gagaaaaaag cTtcaaaaga
      501  agagaaaaag gcggcgaaag aaaAAGttgc acccaaaccg accccggaaC
      551  aaatcctcaa cagccgCagc atcgaaaaag cgcgtagtgc cgctgccaaa
      601  gaAgtgcaGA AAatgaaaaa ctTtgggcaa ggcgGaagcc aacgcattaT
      651  CTGcaaatgg gcgcgtatgc cgaccgtccg gagcgcggaA gggcagcgtg
45      701  ccaaAActggc aAtcttgGgc atatctTccg aagtggtcgG CTATCAGGCG
      751  GGACATAAAA CGCTTTACCG CGTGCAAagc GGCAaatatgt ccgccgatgc
      801  gGTGAAAAAA ATGCAGGACG AGTTGAAAAA GCATGGGGtt gcCAGCCTGA
      851  TCCGTGcgAT TGAAGGCAAA TAA

```

This encodes the following amino acid sequence <SEQ ID 390>:

```

50      1  MFMNKFSQSG KGLSGFFFGL ILATVIIAGI LLYLNQGGQN AFKIPAPSKQ
      51  PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV
      101  ADKADEVEEK AGEPEREEDP QAVRKKALT EEREQTVREK AQKKDAETVK
      151  KQAVKPSKET EKKASKEEK AAKEKVAPKP TPEQILNSRS IEKARSAAAK
      201  EVQKMKNFGQ GGSQRIICKW ARMPTVRSAG QRAKLAILG ISSEVVGYQA
55      251  GHKTLRVQS GNMSADAVKK MQDELKKHGV ASLIRAIEGK *

```

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

```

      10      20      30      40      50      60
orf65-1.pep MFMNKFSQSGKGLSGFFFGLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60      orf65ng-1 MFMNKFSQSGKGLSGFFFGLILATVIIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLK
      10      20      30      40      50      60

```

-246-

		70	80	90	100	110	120
	orf65-1.pep	NQPKEDIQPEPADQNALSEPDAAEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEPD					
5	orf65ng-1	NQPKEDIQPEPADQNALSEPDVAKEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEPD					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf65-1.pep	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETTEKKASKEEKAAKEKVAPKP					
10	orf65ng-1	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETTEKKASKEEKAAKEKVAPKP					
		130	140	150	160	170	180
		190	200	210	220	230	239
15	orf65-1.pep	TPEQILNSGSIEKARSAAAKEVQKMKTSKAEATHYL-QMGAYADRQSAEGORAKLAILG					
	orf65ng-1	TPEQILNSRSIEKARSAAAKEVQKMNFGQGSQRIICKWARMPTVRSAGEGORAKLAILG					
		190	200	210	220	230	240
		240	250	260	270	280	290
20	orf65-1.pep	ISSKVVGQAGHKTLRYVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
	orf65ng-1	ISSEVVGQAGHKTLRYVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX					
		250	260	270	280	290	

- 25 On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

30 391>:

	1	ATGAACCACG	ACATCACTTT	CCTCACCTG	TTCCTACTCG	GTkTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GcGTTTGS.s
	101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
35	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAaATCGGCA	AACCGATATG
	351	GCGGAACCTG	AACCCGATAC	TCAACCGGCT	GTTACCCATA	AAATCCATAC
40	401	CCGCCTGCCT	tGCGgTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGCG	CGTCGCTTTA	CGCGCTGGGA	AgCGGTAGTG	CGGCAACGGG
	501	CGGTTATAT	ATGCTTGCCCT	TTGCACTGGG	TACGCTGCCC	AATCTTtTAG
	551	CAATCGGCAT	TTTtTCCCTG	CAACTGAaWA	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
	651	TGCCGTCCTG	TGGCTGTAA			

- 45 This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

	1	MNHDITFLTL	FLLGXFGGTH	CIGMCGGLSS	AFXXQLPPhi	NRFWLILLLN
	51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
	101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPI	KSIPACLAVG	ILWGWLPcGL
50	151	VYSASLYALG	SGSAATGGLY	MLAFALGTLp	NLLAIGIFSL	QLXKIMQnRY
	201	IRLCTGLSVS	LWALWKLAVL	WL*		

Further work elaborated the DNA sequence <SEQ ID 393> as:

	1	ATGAACCACG	ACATCACTTT	CCTCACCTG	TTCCTACTCG	GTTTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
55	101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG

-247-

5
351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
401 CCGCCTGCCT TGC GGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CCGCAACGGG
501 CGGGTTATAT ATGCTTGCCCT TTGCACTGGG TACGCTGCC AATCTTTAG
551 CAATCGGCAT TTTTCCCTG CAACTGAAAA AAATCATGCA AAACCGATAT
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
651 TGCCGTCCTG TGGCTGTAA

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

10
1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NRFWLILLN
51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVL QNILYTAANL LLLFLGLYLS
101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VG ILWGWLP CGL
151 VYSASLYALG SGSAAATGGLY MLAFALGTLP NLLAIGIFSL QLKIMQNR
201 IRLCTGLSVS LWALWKLAVL WL*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
20	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPhINRFWLILLNLTGRVSSYTAI					
	orf103a	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhINRXWLILLNLTGRVSSYTAI					
		10	20	30	40	50	60
25	orf103.pep	70	80	90	100	110	120
	orf103a	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
30	orf103.pep	130	140	150	160	170	180
	orf103a	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
35	orf103.pep	190	200	210	220		
	orf103a	NLLAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
40		190	200	210	220		
		NLXAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWLX					

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

45
55
1 ATGAACCANG ACATCACTTT CCTCACCTG TTCCTACTCG GTTTCTTCGG
51 CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGC GC
101 TCCAAC TCCC CCCGCATATC AACCGCTTNT GGCTGATCCT GCTGCTTAAC
151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCCTGATAC TCGGATTAAT
201 CGGACAGGTC GCGGTTTCAC TCGACCAAAC CCGCGTCNTG CAGAATATT
251 TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
301 GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAAATCGGCA AACCGATATG
351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
401 CCGCCTGCCT TGC GGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTA
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CCGCAACGGG
501 CGGGTTATAT ATGCTTGCCCT TTGCACTGGG TACGCTGCC AATCTTTNGG
551 CAATCGGCAT TTTTCCCTG CAACTGNAAA AAATCATGCA AAACCGATAT
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
651 TGCCGTCCTG TGGCTGTAA

This encodes a protein having amino acid sequence <SEQ ID 396>:

1 MNXDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NRXLILLN
51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVX QNILYTAANL LLLFLGLYLS
101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VG ILWGWLP CGL

151 VYSASLYALG SGSAATGGLY MLAFALGTLP NLXAIGIFSL QLXKIMQNRY
 201 IRLCTGLSVS LWALWKLAVL WL*

ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

		10	20	30	40	50	60
5	orf103a.pep	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLILLNTGRVSSYTAI					
	orf103-1	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI					
		10	20	30	40	50	60
10	orf103a.pep	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
	orf103-1	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
15	orf103a.pep	NPILNRLLPKISIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
	orf103-1	NPILNRLLPKISIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
20	orf103a.pep	NLXAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
	orf103-1	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		
25	orf103a.pep	NLXAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
	orf103-1	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

30 *gonorrhoeae*:

	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNTGRVSSYTAI	60
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI	60
35	orf103.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
	orf103ng	GLMLGLIGQLGISL DQTRVLQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
40	orf103.pep	NPILNRLLPKISIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP	180
	orf103ng	NPILNRLLPKISIPACLA VGILWGWLP CGLVYSASLYALGSGSATTGGLYMLAFALGTLP	180
	orf103.pep	NLLAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWL	222
45	orf103ng	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

	1	ATGAACCACG	ACATCACTTT	CCTCACCTCG	TTCCTGCTCG	GTTTCTTCGG
	51	CGGAATCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
50	101	TCCAATCCC	CCCGCATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
	151	ACAGGACGGA	TAAGCAGCTA	TACGGCAATC	GGCCTGATGC	TCGGATTAAT
	201	CGGACAATC	GGCATTTCAC	TCGACCAAAc	ccgcgTCCTG	CAAAATATTT
	251	tatacacagc	ctccaaCCTC	CTGCTGCTCT	TTTAGGCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
	351	GCGCAACCTG	AACCCGATAC	TCAACCGGCT	GCTGCCCAT	AAATCCATAC
55	401	CGGCCTGCCT	TGCTGTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGCG	CATCACTTTA	CGCGCTGGGA	AGCGGTAGTG	CGACAACCGG
	501	CGGACTGTAT	ATGCTTGCCCT	TTGCACTGGG	TACGCTGCCC	AATCTTTTGG
	551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACAGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
60	651	TGCCGTCCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

-249-

1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NRFWLILLLN
 51 TGRISSYTAI GLMLGLIGQL GISLDQTRVL QNILYTASNL LLLFLGLYLS
 101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACLAVG ILWGWLPCGL
 151 VYSASLYALG SGSATTGGLY MLAFALGTLF NLLAIGIFSL QLKKIMQNRY
 201 IRLCTGLSVS LWALWKLAVL WL*

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

		10	20	30	40	50	60
10	orf103-1.pep	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhiNRFWLILLLN	TGRVSSYTAI				
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhiNRFWLILLLN	TGRVSSYTAI				
		10	20	30	40	50	60
15	orf103-1.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTASNLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
20	orf103-1.pep	NPILNRLLPIKSIPACLAVGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLF					
	orf103ng	NPILNRLLPIKSIPACLAVGILWGWLPCGLVYSASLYALGSGSATTGGLYMLAFALGTLF					
		130	140	150	160	170	180
25	orf103-1.pep	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
	orf103ng	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

30 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

	1	ATGGA	AAACC	AAAGGCCGCT	CCTAGGCTTT	CGCTTGGCAC	TTTTGGCGGC
	51	GATGACGTGG	GGAACGCTGC	CGAT.TCCGT	GCGGCAGGTA	TTGAAGTTTG	
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA	
40	151	TTGTTTGT	TGCTGGCACT	GGCGGGCGG	CTGCcGAAGC	GGCGaGGATT	
	201	TTTCTTGGTG	CTCATTCAAG	CTGCTGCTGC	TCGGCGTGGC	GGGCATTTCG	
	251	GCAAACCTTTG	TGCTGATTGC	CCAAGGGCTG	CATTATATTT	CGCCGACCAC	
	301	GACGCAGGTT	TTGTGGCAGA	TTTCGCCGTT	TACGATGATT	GTWGTCCGTG	
	351	TGTTGGTGTT	TAAAGACCGG	ATGACTGCCG	CTCAGAAAAT	CGGCTTGGTT	
45	401	TTGCTGCTTG	CCGGTTTGCT	TATGTATTTT	AACGATAAAT	TCGGCGAGTT	
	451	GTCGGGTTTG	GGCGCGTATG	C.AAGGGCGT	GTTGCTGTGT	GCGGCAGGCA	
	501	GTATGGCATG	GGTGTGTAAT	GCCGTGGCGC	AAAAGCTGCT	GTCGGCGCAA	
	551	TTCCGGCCGC	AACAGATTCT	GCTGTTGATT	TATGCGGCAA	GTGCGCGCCT	
	601	GTTCTGCCG	TTTGCCGAAC	CGGCACACAT	CGGAAGTATG	GACGGTACGT	
	651	TGGCGTGGGT	ATGTATTGCG	TATTGCTGCT	TGAATACGTT	AATCGGTTAC	
50	701	GGCTCGTTTC	GCGAGGCGTT	GAAACATTGG	GAGGCTTCCA	AAGTCAGCGC	
	751	GGTAACAACC	TTGCTCCCGG	TGTTACCCTG	AATAAATACT	TTGCTCGGGC	
	801	ATTATGTGAT	GCCTGAAACT	TTTGCCGCGC	CGGA..		

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

55	1	MENQR	PLLGF	RLALLAAMTW	GTLPXSVRQV	LKFVDAPTLV	WVRFTVAAAV
	51	LFVLLALGGR	LPKRRDFSWC	SFRLLLLGVA	GISANFVLIA	QGLHYISPTT	
	101	TQVLWQISPF	TMIVGVVLVF	KDRMTAAQKI	GLVLLLAGLL	MYFNDKFGEI	
	151	SGLGAYXKGV	LLCAAGSMAW	VCNAVAQKLL	SAQFGPQQIL	LLIYAASAAV	
	201	FLPFAEPAHI	GSMDGTLAWV	CIAYCCLNTL	IGYGSFGEAL	KHWEASKVSA	

251 VTLLPVFTV INTLLGHYVM PETFAAP...

Further work revealed further partial DNA sequence <SEQ ID 401>:

```

      1 ATGGAACACC AAAGCCGCT CCTAGGCTTC GCGTTGGCAC TTTTGGCGGC
      51 GATGACGTGG GGAACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
5   101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GGCGGCGGTA
      151 TTGTTTGTGT TGCTGGCACT GGGCGGGCGG CTGCCGAAGC GGCGGGATTT
      201 TTCTTGGTGC TCATTCAGGC TGCTGCTGCT CGGCGTGGCG GGCATTTCCG
      251 CAAACTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTTT GCCGACCACG
10  301 ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATTG TTGTCGGTGT
      351 GTTGGTGTGT AAAGACCGGA TGACTGCCGC TCAGAAAATC GGCTTGGTTT
      401 TGCTGCTTGC CGGTTTGCTT ATGTTTTTTA ACGATAAATT CGGCGAGTTG
      451 TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CGGCAGGCAG
      501 TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCGCAAT
      551 TCGGGCCGCA ACAGATTCTG CTGTTGATTT ATGCGGCAAG TGCCGCCGCTG
15  601 TTCTGCCGT TTGCCGAACC GGCACACATC GGAAGTTTGG ACGGTACGTT
      651 GCGGTGGGTT TGTTTTGCCT ATTGCTGCTT GAATACGTTA ATCGGTTACG
      701 GTCGTTCCG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
      751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATAWTWCTT TGCTCGGGCA
      801 TTATGTGATG CCTGAAACTT TTGCCGCGCC GGA...

```

20 This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>:

```

      1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
      51 LFVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPTT
      101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL
      151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
25  201 FLPPFAEPAHI GSLDGT LAWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
      251 VTLLPVFTV IXXLLGHYVM PETFAAP...

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

```

30  orf104  4  QRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWVXXXXXXXXXXXXXXXXXXXXP- 62
      Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P
      HI0878  3  QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

      orf104  63  --KRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 120
      K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
35  HI0878  63  LMKVRQYAW----IMLIGVIGLTSNFLFLSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118

      orf104  121 KDRMTAAQKIXXXXXXXXXXXMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180
      K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+
40  HI0878  119 KEKLGHLHQIGLFLLLIGLGLFFNDRFADFAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178

      orf104  181 SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL 240
      +F QQILL++Y A F+P A+ + + + LA +C YCCLNTLIGYGS+ EAL
45  HI0878  179 LRKFNSQQIILLMYLGCAIAFMPMADFSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237

      orf104  241 KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277
      W+ SKVS V TL+P+FT++ + + HY P FAAP
      HI0878  238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAP 274

```

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N.*

meningitidis:

```

55  orf104.pep 10 20 30 40 50 60
      MENQRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWVRFVAAAVLFVLLALGGR
      ||||| : |||||
      orf104a 10 20 30 40 50 60
      MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFVAAAVLFVLLALGGR
      70 80 90 100 110 120

```


-251-

5	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF
	orf104a	LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF
10	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL
	orf104a	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
15	orf104.pep	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL
	orf104a	SAQFGPQQILLIYAASAAVFLPFAELAHIGSLDGTALWVCFAYCCLNTLIGYGSFGEAL
20	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP
	orf104a	KHWEASKVSAVTTLLPVFTVIFSLLLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

1	ATGGA	AAACC	AAAGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTTGGCGGC
25	51	GATGACGTGG	GGAAACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA
	151	TTGTTTGT	TGCTGGCATT	GGGCGGGCGG	CTGCCGAAGT	GGCGGGATT
	201	TTCTTGGTGC	TCATTACAGGC	TGCTGCTGCT	CGGCGTGGCG	GGCATTTCGG
30	251	CAAACTTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTC	GCCGACCACG
	301	ACGCAGGTTT	TGTGGCAGAT	TCGCCGTTT	ACGATGATTG	TTGTCCGGTG
	351	GTTGGTGT	AAAGACCGGA	TGACTGCCGC	TCAGAAAATC	GGCTTGGTTT
	401	TGCTGCTTGC	CGGTTTGCTT	ATGTTTTTTA	ACGATAAATT	CGGCGAGTTG
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
35	501	TATGGCATGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCGCAAT
	551	TCGGGCCGCA	ACAGATTCTG	CTGTTGATT	ATGCGGCAAG	TGCCGCCGTG
	601	TTCTTGCCGT	TTGCCGAAT	GGCACACATC	GGAAGTTTGG	ACGGTACGTT
	651	GGCGTGGGTT	TGTTTTCGCT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG
	701	GCTCGTTTCGG	CGAGGCGTTG	AAACATTGGG	AGGCTTCCAA	AGTCAGCGCG
40	751	GTAACAACCT	TGCTCCCCGT	GTTTACCGTA	ATATTTTCTT	TGCTCGGGCA
	801	TTATGTGATG	CCTGATACTT	TTGCCGCGCC	GGATATGAAC	GGTTTGGGTT
	851	ATGCCGGCGC	ACTGTCGTG	GTCCGGGGTG	CGGTTACGGC	GGCGGTGGGG
	901	GACAGGCTGT	TCAAACGCCG	CTAG		

This encodes a protein having amino acid sequence <SEQ ID 404>:

1	MENQR	PLLGF	ALALLA	AMTW	GTLP	IAVRQV	LKFVD	APTLV	WVRFT	VAAAV
45	51	LFVLL	ALGGR	LPKWR	DFSWC	SFRLL	LLGVA	GISAN	FVLIA	QGLHY
	101	TOVLW	QISPF	TMIVV	GVLVF	KDRMT	AAQKI	GLVLL	LAGLL	MFFND
	151	SLGAY	AKGV	LLCAAG	SMAW	VCYAV	AQKLL	SAQFG	PQQIL	LLIYA
	201	FLPFA	ELAH	IGSLD	GTALW	CFAYC	CLNTL	IGYGS	FGEAL	KHWEA
50	251	VTTLL	VPVFTV	IFSL	LGHYVM	PDTFA	APDMN	GLGYA	GALVV	VGGAV
	301	DRLE	FKRR*							

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

55	orf104a.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFVLLALGGR
	orf104-1	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFVLLALGGR
60	orf104a.pep	LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF
	orf104-1	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF
65	orf104a.pep	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL

[illegible]

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orf104.pep	MENQRPLLGFRLALLAAMTWGTLPSXSVRQVLKFVDAPTLVWVRFTVAAAVLVFLLALGGR	60
	orf104ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLVFLLALGGR	60
25	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF	120
	orf104ng	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF	120
30	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLCAAGSMAWVCNAVAQKLL	180
	orf104ng	KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
35	orf104.pep	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL	240
	orf104ng	SAQFGPQQIILLIYAASAAVFLXAEPAHIGSLDGTALWVCFVYCCLNTLIGYGSFGEAL	240
	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP	277
	orf104ng	KHWEASKVSAVTTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGYVGALVVGGAVTAAVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a
40 protein having amino acid sequence <SEQ ID 406>:

	1	<u>MENQRPLLGF</u>	<u>ALALLAAMTW</u>	GTLPiAVRQV	LKFVDAPTLV	WVRFTVAAAV
	51	<u>LFVLLALGGR</u>	<u>LPKRDRFSWH</u>	SFRLLLLGVT	GISANFLVIA	QGLHYISPTT
	101	<u>TOVLWQISPF</u>	<u>TMIVVGVLVF</u>	KDRMTAAQKI	<u>GLVLLVLGLL</u>	MFNDKFGEL
	151	<u>SGLGAYAKGV</u>	<u>LLCAAGSMAW</u>	VCYAYAQKLL	<u>SAQFGQOIL</u>	LLFYAASAAV
45	201	<u>FLLXAEPAHI</u>	<u>GSLDGTIAWV</u>	CFVYCCLNTL	IGYGSFGEAL	KHWEASKVSA
	251	<u>VTTLVPVFTV</u>	<u>IFSLLGHYVM</u>	PDTFAAPDMN	<u>GLGYVGALVV</u>	VGGAVTAAVG
	301	<u>DRPFKRR*</u>				

50	1	ATGGA AAAACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTTGGCGGC
	51	GATGACGTGG	GGGACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTAGC	GGCGGCGGTA
	151	TTGTTTGGTT	TGCTGGCATT	GGCGGGGCGG	TGCCGAAGC	GGCGGGATT
	201	TTCTTGGCAT	TCATTACAGC	TGCTGCTGCT	CGGCGTAGC	GGCATTTCGG
55	251	CAAACTTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTT	GCCGACCACG
	301	ACGAGGTTT	TGTGGCAGAT	TTCGCCGTTT	ACGATGATT	TTGTTCGGCT
	351	GTTGGTGT	AAAGACCGGA	tgaCTGCCGC	GCAGAAAATC	GGTTTGGTTT
	401	TGCTGCTtgT	CGGTTtgCTT	ATGTTTTtta	ACGACAAAT	CGCGCAGTTG
	451	TCGGGTTTGT	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
60	501	TATGGCCTGG	GTGTGTTATG	CCGTGGCGTA	AAAGCTGCTG	TCGGCGCAAT
	551	TCGGGCCCGA	ACAGATTCTG	CTGTTGATT	ATGCGGCaag	tgcgcgCTGT
	601	TTCTgtgcgT	TTGccgaaCC	GGCACACATC	GGAAGTTTgg	aCGTTACGtt
	651	GGCGTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG

701 GCTCGTTCGG CGAGGCGTTG AACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATATTTCTT TGCTCGGGCA
 801 TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTTTGGGTT
 851 ATGTCGGCGC ACTGGTCGTG GTCGGGGGTG CGGTTACGGC GCGGTGGGG
 901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT
 101 TOVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLVGLL MFFNDKFGEL
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
 201 FLPPAEPahi GSLDGT LAWV CFVYCCLNTL IGYGSFGEAL KHWEASKVSA
 251 VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG
 301 DRPFKRR*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15		10	20	30	40	50	60
	orf104-1.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFVLLALGGR					
	orf104ng-1	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFVLLALGGR					
20		10	20	30	40	50	60
	orf104-1.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF					
	orf104ng-1	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF					
25		70	80	90	100	110	120
	orf104-1.pep	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
	orf104ng-1	KDRMTAAQKIGLVLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
30		130	140	150	160	170	180
	orf104-1.pep	SAQFGPQQIILLIYAASAAVFLPPAEPahIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL					
	orf104ng-1	SAQFGPQQIILLIYAASAAVFLPPAEPahIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL					
35		190	200	210	220	230	240
	orf104-1.pep	KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP					
	orf104ng-1	KHWEASKVSAVTTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG					
40		250	260	270	280	290	300

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45 gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306
 Score = 237 bits (598), Expect = 8e-62
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

50 Query: 30 QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 88
 Q+P M WG+LPPIA++QVL ++A T+VW P
 Sbjct: 3 QPPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

55 Query: 89 --KRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 146
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
 Sbjct: 63 LMKVRQYAW----IMLIGVIGLTSNFLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118

60 Query: 147 KDRMTAAQKIXXXXXXXXXXXMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206
 K+++ QKI +FFND+F +GL Y+ GV+L G++ WV Y +AQKL+
 Sbjct: 119 KEKLGHLHQIGLFLLLIGLGLFFNDRDFAAGLNQYSTGVILGVGALIVVAYGMAQKLM 178

65 Query: 207 SAQFGPQQIILLIYAASAAVFLPPAEPahIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL 266
 +F QQILL++Y A F+P A+ + + L LA +CF+YCCLNTLIGYGS+ EAL
 Sbjct: 179 LRKFNSQQILLMYLGCAIAFMPMADFQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLGHYVMPDTFAAPDMN 306

W+ SKVS V TL+P+FT++FS + HY P FAAP++N
 Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from
 5 *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

10      1  ATGGTAGCTC  GTCGGGCTCA  TAACCCGAAG  GTCGTAGGTT  CGAATCCTGT
      51  .CCC GCAACC  TAATTTCAAA  CCCCTCGGTT  CAATGCCGAG  GG.GTTTGT
     101  T.TTGCCTGT  TTCCTGTTTC  CTGTTTCCTG  CCGCCTCCGT  TTTTGTCCGG
     151  ATTTTCCTTC  CGGCCGCAAT  ATCGGAACGG  CAGACCGCCG  TCTGTTTGCG
     201  GTTGCAAATT  CAGGCAGTTT  GGCTACAATC  TTCCGCATTG  TCTTCAAGAA
     251  AGCCAACCAT  GCCGACCGTC  CGTTTACCG  AATCCGTCAG  CAAACAAGAC
    15  301  CTTGATGCTC  TGTTCGAGTG  GGCAAAAGCA  AGTTACGGTG  CAGAAAGTTG
     351  CTGGAAAACG  CTGTATCTGA  ACGGTCysCC  TTTGGGCAAC  CTGTCGCCGG
     401  AATGGGTGGA  ACGCGTsmmA  AAAGACTGGG  AGGCAGGCTG  CyCGGAGTCT
     451  TCAGACGGCA  TTTTCTGAA  TgCGGACGGc  TGgCctGATA  TGGgCGGAcg
     501  cTTACAGCAC  CTCGCCCTCG  GTTGGCACTG  TGCGGGGCTG  TTGGACGgsT
    20  551  GGCGCAACGA  GTGTTTCGAC  CTGACCGACG  GCGGCGGCAA  CCCCTTGTTT
     601  ACGCTCGaAc  GCGCCGyTTT  mCGTCCTkTC  GGACTGCTCA  GCCGCGCCGT
     651  CCATCTCAAC  GGTCTGACCG  AATCGGACGG  CCGATGGCAT  TTCTGGATAG
     701  GCAGGCGCAG  TCCGCACAAA  GCAGTCGATC  CCAACAAACT  CGACAATACT
     751  rCCGCCGGCG  GTGTTTCCGG  CGGCGAAATG  CCGTCTGAAG  CCGTGTGTCT
    25  801  CGAAAGCAGC  GAAGAAGCCG  GTTGGATAAA  AACGCTGcTT  CCGCTCATCC
     851  GCCCGGTATC  GCAGCTGCAC  AGCCTGCGCT  CCGTCAGCCG  GGGTGTACAC
     901  AATGAAATCC  TGTATGTATT  CGATGCCGTC  CTGCCG...
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

30      1  MVARRAHNP  VVGSNPXPAT  XFQTPRFNAE  XVLXLPVSCF  LFPAASVFCR
     51  IFLPAAISER  QTAVCLRLQI  QAVWLQSSAL  SSRKPTMPTV  RFTESVSKQD
     101  LDALFEWAKA  SYGAESCWKT  LYLNGXPLGN  LSPewVERVX  KDWEAGCXES
     151  SDGIFLNADG  WPDmGGRlQH  LALGWHCAGL  LDGWRNECFD  LTDGGGNPLF
     201  TLERAXRXPX  GLLSRVHLN  GLTESDGRWH  FWIGRRSPHK  AVDPNKLDNT
     251  XAGGVSGGEM  PSEAVCRESS  EEAGLDKTL  PLIRPVSQLH  SLRSVSRGVH
    35  301  NEILYVFDAV  LP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

40      1  ATGCCGACCG  TCCGTTTTAC  CGAATCCGTC  AGCAAACAAG  ACCTTGATGC
     51  TCTGTTCGAG  TGGGCAAAAG  CAAGTTACGG  TGCAGAAAGT  TGCTGGAAAA
     101  CGCTGTATCT  GAACGGTCTG  CCTTTGGGCA  ACCTGTGCGC  GGAATGGGTG
     151  GAACGCGTCA  AAAAAGACTG  GGAGGCAGGC  TGCTCGGAGT  CTTCAGACGG
     201  CATTTTTCTG  AATGCGGACG  GCTGGCCTGA  TATGGGCGGA  CGCTTACAGC
     251  ACCTCGCCCT  CGGTTGGCAC  TGTGCGGGGC  TGTGGACGG  CTGGCGCAAC
     301  GAGTGTTCG  ACCTGACCGA  CGGCGGCGGC  AACCCTTGT  TCACGCTCGA
     351  ACGCGCCGCT  TTCCGTCTCT  TCGGACTGCT  CAGCCGCGCC  GTCCATCTCA
    45  401  ACGGTCTGAC  CGAATCGGAC  GGCCGATGGC  ATTTCTGGAT  AGGCAGGCGC
     451  AGTCCGCACA  AAGCAGTCGA  TCCCAACAAA  CTCGACAATA  CTGCCGCCGG
     501  CCGTGTTC  GCGGCGGAAA  TGCCGTCTGA  AGCCGTGTGT  CGCGAAAGCA
     551  GCGAAGAAGC  CGGTTTGGAT  AAAACGCTGC  TTCCGCTCAT  CCGCCCGGTA
     601  TCGCAGCTGC  ACAGCCTGCG  CTCCGTGAGC  CGGGGTGTAC  ACAATGAAAT
    50  651  CCTGTATGTA  TTGATGCGG  TCCTGCCCGA  AACCTTCTG  CCTGAAAATC
     701  AGGATGGCGA  AGTGGCGGGT  TTTGAGAAAA  TGGACATCGG  CGGTCTGTTG
     751  GATGCCATGT  TGTGCGGAAA  CATGATGCAC  GACGCGCAAC  TGGTTACGCT
     801  GGACGCGTTT  TGCCGTTACG  GTCTGATTGA  TGCCGCCCAT  CCGCTGTCCG
     851  AGTGCTGGA  CGGCATACGT  TTATAG
```

55 This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

-255-

1 MPTVRFTESV SKQDLDALE WAKASYGAES CWKTLYLNGL PLGNLSPEWV
 51 ERVKKDWEAG CSESSDGIFL NADGWPDMMG RLQHLALGWH CAGLLDGWRN
 101 ECFDLTDGGG NPLFTLERAA FRPFGLLSRA VHLNGLTESD GRWHFWIGRR
 151 SPHKAVDPNK LDNTAAGGVS GGEMPSEAVC RESSEEAGLD KTLPLIRPV
 201 SQLHSLRSVS RGVHNEILYV FDAVLBPETFL PENQDGEVAG FEKMDIGGLL
 251 DAMLSGNMMH DAQLVTLDAF CRYGLIDAAH PLSEWLDGIR L*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

10 *meningitidis*:

	60	70	80	90	100	110
orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAES					
orf105a	MPTVRFTESVSKHDLDALEWAKASYGAES					
				10	20	30
orf105.pep	120	130	140	150	160	170
orf105a	CWKTLYLNGXPLGNLSPEWVERVXKDWEAGCXESSDGIFLNADGWPDMMGRLQHLALGWH					
	40	50	60	70	80	90
orf105.pep	180	190	200	210	220	230
orf105a	CAGLLDGWRNECFDLTDGGGNPLFTLERAXRXPGLLSRAVHLNGLTESDGRWHFWIGRR					
	100	110	120	130	140	150
orf105.pep	240	250	260	270	280	290
orf105a	SPHKAVDPNKLDNTXAGGVSSGGEMPSEAVCRESSEEAGLDKTLPLIRPVSQHLHSLRSVS					
	160	170	180	190	200	210
orf105.pep	300	310				
orf105a	RGVHNEILYVFDVLPEFLPENQDGEVAGFEKMDIGGLLAAMLSGNMMHDAQLVTLDAF					
	220	230	240	250	260	270

40 The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

1 ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACACG ACCTTGATGC
 51 CCTATTCGAG TGGGCAAAGG CAAGTTACGG TGCGGAAAGT TGCTGGAAAA
 101 CGCTGTATCT GAACGGTCTG CCTTTGGGCA ATCTGTCGCC GGAATGGGCG
 151 GAGCGCGTCA AAAAAGACTG GGAGGCAGGC TGCTCGGAGT CTTACAGACG
 201 CATTTTCCTG AATGCGGACG GCTGGCCAGA TATGGGCAGA CGCTTGACAGC
 251 ACCTCGCCCG AATATGGAAA GAAGCGGGAC TGCTTCACGG CTGGCGCGAC
 301 GAGTGTTCG ACCTGACCGA CGGCGGCAGC AATCCCTTGT TCGCGCTCGA
 351 ACGCGCGGCT TTCCGTCCGT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
 401 ACGGTTTGGT CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAGGCGC
 451 AGTCCGCACA AAGCAGTCGA TCCCGACAAA CTCGACAATA CTGCCGCCGG
 501 CGGTGTTTCC AGCGGTGAAT TGCCGTCTGA AACCGTGTGT CGCGAAAGCA
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCCGGTA
 601 TCGCAGCTGC ACAGCCTGCG CCCCCTCAGC CGGGGTGTGC ACAATGAAAT
 651 CCTGTATGTA TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCTGAAATC
 701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CGGTCTGTG
 751 TGTCGCATGT TGTCGGGAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
 801 GGACGCGTTT TGCCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This encodes a protein having amino acid sequence <SEQ ID 414>:

60 1 MPTVRFTESV SKHDLDALE WAKASYGAES CWKTLYLNGL PLGNLSPEWA
 51 ERVKKDWEAG CSESSDGIFL NADGWPDMMGR RLQHLARIWK EAGLLHGWDR

-256-

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR
 151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV
 201 SQLHSLRPVS RGVHNEILYV FDAVLPEFL PENQDGEVAG FEKMDIGLL
 251 AAMLSGNMMH DAQLVTDAF CRYGLIDAAH PLSEWLDGIR L*

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

		10	20	30	40	50	60
	orf105a.pep	MPTVRFTESVSKHDL	DALFEWAKASYGAESC	WKTLYLNGPLGNLS	PEWAERVKKDWEAG		
	orf105-1	MPTVRFTESVSKQDL	DALFEWAKASYGAESC	WKTLYLNGPLGNLS	PEWVERVKKDWEAG		
10		10	20	30	40	50	60
		70	80	90	100	110	120
	orf105a.pep	CSESSDGIFLNADGW	PDMGRRLOHLARIWKE	AGLLHGWRDEC	FDLTDGGSNPLFALERAA		
	orf105-1	CSESSDGIFLNADGW	PDMGRRLOHLALGWHC	AGLLDGWRNECFDLTDGGSNPLFTLERAA			
15		70	80	90	100	110	120
		130	140	150	160	170	180
	orf105a.pep	FRPFGLLSRAVHLNGL	VESDGRWHFWIGRRSPHK	AVDPDKLDNTAAGGVSSGELPSETVC			
	orf105-1	FRPFGLLSRAVHLNGL	TESDGRWHFWIGRRSPHK	AVDPKNKLDNTAAGGVSSGEMPSEAVC			
20		130	140	150	160	170	180
		190	200	210	220	230	240
	orf105a.pep	RESSEEAGLDKTLPLIR	PVSQHLHSLRPVSRGVHNEILYV	FDAVLPEFLPENQDGEVAG			
	orf105-1	RESSEEAGLDKTLPLIR	PVSQHLHSLRSVSRGVHNEILYV	FDAVLPEFLPENQDGEVAG			
25		190	200	210	220	230	240
		250	260	270	280	290	
	orf105a.pep	FEKMDIGLLAAMLSGN	MMHDAQLVTDAFCRYGLIDAAH	PLSEWLDGIRLX			
	orf105-1	FEKMDIGLLDAMLSGN	MMHDAQLVTDAFCRYGLIDAAH	PLSEWLDGIRLX			
30		250	260	270	280	290	
35							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N. gonorrhoeae*:

40	orf105.pep	MVARRAHNPKVVGSNPXPATXFQTPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER	60
	orf105ng	MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER	55
	orf105.pep	QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	120
45	orf105ng	QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	115
	orf105.pep	LYLNGXPLGNLSPEWVERVXKDWEAGCXESSDGIFLNADGWPDMMGRRLOHLALGWHCAGL	180
50	orf105ng	LYLNRLPLGNLSPEWAERIKKDWEAGCSESSNGIFLNADGWPDMMGRRLOHLARTWNKAGL	175
	orf105.pep	LDGWRNECFDLTDGGSNPLFTLERAXRXPXGLLSRAVHLNGLTESDGRWHFWIGRRSPHK	240
	orf105ng	LHGWRNECFDLTDGGSNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK	235
55	orf105.pep	AVDPNKLDNTXAGGVSGGEMPSEAVCRESSEEAGLDKTLPLIRPVSQHLHSLRSVSRGVH	300
	orf105ng	AVDPGKLDNIAGGVSGGEMPSEAVCRESSEEAGLDKTLPLIRPVSRLHSLRPVSRGVH	295
60	orf105.pep	NEILYVFDVLP	312
	orf105ng	NEILYVFDVLPETFLPENQDGEVAGFEKMDIGLLDAMLSKNMMHDAQLVTDAFYRYG	355

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

-257-

1 MVARRAHNPK VVGSNPAPAT KYQTPRENAE GVLFFLFPPAA SVFCRIFLPA
 51 AISERQAAVC LRLQIQAVWL QSSALCSRKP AMPTVRFTES VSKQDLDALE
 101 ERAKASYGAE SCWKTLYLNR LPLGNLSPEW AERIKKDWEA GCSESSNGIF
 151 LNADGWPDMD GRLQHLARTW NKAGLLHGWR NECFDLTDGG GNPLFTLERA
 5 201 AFRPFGLLIR AVHLNGLVES NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV
 251 SGGEMPSEAV CRESSEEAGL DKTFLPLIRP VSRLHSLRPV SRGVHNEILY
 301 VFDAVLPETF LPENQDGEVA GF EKMDIGGL LDAMLSKNMM HDAQLVTLDA
 351 FYRYGLIDAA HPLSEWLDGI RL*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

10 1 ATGCCGACCG TCCGTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC
 51 CCTGTTTCGAG CGGGCAAAAG CAAGTTACGG TGCCGAAAGT TGCTGGAAAA
 101 CGCTGTATCT GAACCGTCTT CCTTTGGGCA ATCTGTGCGC GGAATGGGCT
 151 GAGCGCATCA AAAAAGACTG GGAGGCAGGC TGCTCCGAGT CTTACAGACGG
 201 CATTTTTCTG AATGCCGACG GCTGGCCGGA TATGGCCGGA CGCTTGACGC
 15 251 ACCTCGCCCG CACATGGAAC AAGGCGGGGC TGCTTCACGG ATGGCGCAAC
 301 GAGTGTTCG ACCTGACCGA CGGCGGCGGC AACCCTTGT TCACGCTCGA
 351 ACGCGCCGCT TTCCGTCGCT TCGGACTACT CAGCCGCGCC GTCCATCTCA
 401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC
 451 AGTCCGCACA AAGCAGTCGA TCCCGGCAAG CTCGACAATA TTGCCGCGCG
 20 501 CGGTGTTTCC GCGCGCGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCAGTA
 601 TCGCGGCTGC ACAGCCTTCG CCCCCTCAGC CGAGGTGTGC ACAATGAAAT
 651 CCTGTATGTG TTCGATGCCG TCCTGCCCGA AACCTTCTG CCTGAAATC
 701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG
 25 751 GATGCCATGT TGTCGAAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
 801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

30 1 MPTVRFTESV SKQDLDALE RAKASYGAES CWKTLYLNRL PLGNLSPEWA
 51 ERIKKDWEAG CSESSDGIFL NADGWPDMDG RLQHLARTWN KAGLLHGWRN
 101 ECFDLTDGGG NPLFTLERA FRPFGLLSRA VHLNGLVESN GRWHFWIGR
 151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEEAGLD KTLFPLIRPV
 201 SRLHSLRPVS RGVHNEILYV FDAVLPETF PENQDGEVAG FEKMDIGGLL
 251 DAMLSKNMMH DAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L*

35 ORF105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
orfl05-1.pep		MPTVRFTESVSKQDLDALE	FEWAKASYGAES	CWKTLYLNRL	PLGNLSPEW	VERVKKDWEAG	
orfl05ng-1		MPTVRFTESVSKQDLDALE	FERAKASYGAES	CWKTLYLNRL	PLGNLSPEW	AERIKKDWEAG	
40		10	20	30	40	50	60
orfl05-1.pep		CSESSDGIFLNADGWPDMD	GRLQHLALGWH	CAGLLDGRNE	CFDLTDGG	GNPLFTLERA	
orfl05ng-1		CSESSDGIFLNADGWPDMD	GRLQHLARTWN	KAGLLHGWRNE	CFDLTDGG	GNPLFTLERA	
45		70	80	90	100	110	120
orfl05-1.pep		FRPFGLLSRAVHLNGL	TESDGRWHFWIGR	RSPHKAVDPN	KLDNTAAG	GVSGGEMPSEAVC	
orfl05ng-1		FRPFGLLSRAVHLNGL	VESNNGRWHFWIGR	RSPHKAVDPG	KLDNIAGG	GVSGGEMPSEAVC	
50		130	140	150	160	170	180
orfl05-1.pep		RESSEEAGLDKTLPL	IRPVSQHL	SLRVS	SRGVHNEILY	VFDAVLPETFL	PENQDGEVAG
orfl05ng-1		RESSEEAGLDKTLPL	IRPVSRLH	SLRVS	SRGVHNEILY	VFDAVLPETFL	PENQDGEVAG
55		190	200	210	220	230	240
orfl05-1.pep		FEKMDIGGLLDAMLS	SGNMMHDAQL	VTLDAFCRY	GLIDAAHPL	SEWLDGIRLX	
orfl05ng-1		FEKMDIGGLLDAMLS	SKNMMHDAQL	VTLDAFYR	YGLIDAAHPL	SEWLDGIRLX	
60		250	260	270	280	290	

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

5  sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
   >gi|1076928|pir|S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
   (Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
   [Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
   pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
   Score = 105 bits (259), Expect = 4e-22
   Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

10 Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAAPFPFGLLSRAVHLNGLVESNGRW--HFWI 441
    N G+ WRNE + + P+ +ER F FG LS VH + + W+
    Sbjct: 96 NTFGIADQWRNELYTVYGKSKKPVLAVERGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155

15 Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESEEAGLDKTLFPLIRPVSRLHSLR 621
    RRSP K P LDN GG++ G+ + +E SEEA LD + LI P + ++
    Sbjct: 156 PRRSPTKQTPWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNLI-PCGTVSYIK 214

20 Query: 622 PVSRG-VHNEILYVFDAVLPEPFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVT 798
    R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
    Sbjct: 215 MEKRHWIQPELQYVFDLVPVDDLVIPIRINDGEVAGFSLPLNQVLHELELKSFKPNCALVL 274

    Query: 799 LDAFYRYGLIDAAHP 843
    LD R+G+I HP
    Sbjct: 275 LDFLIRHGIITPQHP 289

```

25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 30 419>:

```

1  ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
51  CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
151 TTGATATTG GTAACATATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
35 201 ACCTGCATCG GCGGTAATCA GGGTGTATGC ACCGgATACG rGkACAATTA
251 CAGCGAAATT CGTGGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
301 TTTGCGCTTT CGACCTCACG TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGCCAGAA CAGGAAGCTG
40 401 GTCGTCTGAA GCTGATACAC GGGAATGAAA CGCGCAgCcT TAAAGCAACT
451 GTCGAACGTT TGGAAAACCA GGAAGTCCAT ATTTGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45 1  MNRPKQFFFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51 51  LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
101 101 FALSTSFRGA GGSVQQQLKT EAVLKKTLAE QELGRLKLIH GNETRSLKAT
151 151 VERLENQELH ISQQIDGQKR RIRLAEEMLO KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

10 20 30 40 50 60

-259-

```

5      orf107.pep  MNRPKQPFFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT
      orf107a     MNRPKQPFFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT
              10      20      30      40      50      60

10     orf107.pep  TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT
      orf107a     TVEGQILPASGVIRVYAPDGTITAKFXEDGEKVKAGDKLFALSTSRFGAGDSVQQQLKT
              70      80      90      100     110     120

15     orf107.pep  EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO
      orf107a     EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO
              130     140     150     160     170     180

20     orf107.pep  KYRFLSXQX
      orf107a     KYRFLSANDAVPKQEMMNVKAELLEQKAKLDAYRREEVGLLQEIQTQNLTLXSLPQAA
              189
              190     200     210     220     230

```

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

```

25      1  ATGAATAGAC  CCAAGCAACC  NTTCTTCCGT  CCCGAAGTCG  CCGTTGCCCG
      51  CCAAACCAGC  CTGACGGGTA  AAGTGATTCT  GACACGACCG  TTGTCATTTT
     101  CCCTATGGAC  GACATTGCA  TCGATATCTG  CGTTATTGAT  TATCCTGTTT
     151  TTGATATTTG  GTAACATAC  GCGAAAGACA  ACAGTGGAGG  GACAAATTTT
     201  ACCTGCATCG  GCGTAATCA  GGGTGTATGC  ACCGGATACG  GGGACAATTA
     251  CNGCGAAATT  CNTGGAAGAT  GGAGAAAAGG  TTAAGGCTGG  CGACAAGCTA
     301  TTTGCGCTTT  CGACCTCACG  TTTGCGCGCA  GGAGATAGCG  TGCAGCAGCA
     351  GTTGAAAACG  GAGGCAGTTT  TGAAGAAAAC  GTTGGCAGAA  CAGGAACTGG
     401  GTCGTCTGAA  GCTGATACAC  GGGAATGAAA  CGCGCAGCCT  TAAAGCAACT
     451  GTCGAACGTT  TGGAAAACCA  GGAATCCAT  ATTTGCAAC  AGATAGACGG
     501  TCAGAAAAGG  CGCATTAGAC  TTGCGGAAGA  AATGTTGCAG  AAATATCGTT
     551  TCCTATCCGC  CAATGATGCA  GTGCCAAAAC  AAGAAATGAT  GAATGTCAAG
     601  GCAGAGCTTT  TAGAGCAGAA  AGCCAAACTT  GATGCCTACC  GCCGAGAAGA
     651  AGTCGGGCTG  CTTCAGGAAA  TCCGCACGCA  GAATCTGACA  TTGGNNAGCC
     701  TCCCCAAGC  GGCATGA

```

This encodes a protein having amino acid sequence <SEQ ID 422>:

```

40      1  MNRPKQPFFR  PEVAVARQTS  LTGKVILTRP  LSFSLWTTFA  SISALLIILF
      51  LIFGNYTRKT  TVEGQILPAS  GVIRVYAPDT  GTITAKFXED  GEKVKAGDKL
     101  FALSTSRFGA  GDSVQQQLKT  EAVLKKTAE  QELGRLKLIH  GNETRSLKAT
     151  VERLENQELH  ISQQIDGQKR  RIRLAEEMLO  KYRFLSANDA  VPKQEMMNVK
     201  AELLEQKAKL  DAYRREEVGL  LQEIRTQNL  LXSLPQAA*

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N. gonorrhoeae*:

```

50     orf107.pep  MNRPKQPFFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT  60
      orf107ng    MNRPKQPFFRPEVAIARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT  60

      orf107.pep  TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT  120
      orf107ng    TMEGQILPASGVIRVYAPDGTITAKFVEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT  120

55     orf107.pep  EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO  180
      orf107ng    EAVLKKTAEQELGRLKLIHENETRSLKATVERLENQKLHISQQIDGQKRRIRLAEEMLR  180

60     orf107.pep  KYRFLSXQ  188
      orf107ng    KYRFLSAQ  188

```

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

      1  MNRPKQPFRR PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
      51  LIFGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL
5      101  FALSTSRFGA GGSVQQQLKT EAVLKKTLAE QELGRLKLIH ENETRSRKAT
      151  VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ*

```

Based on the presence of a putative ransmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

      1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
      51  GTGCGGCAAA TCCGTAAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA
15     101  GCGCGCCGAA ACCGGTTTTT AAAGTCATAT ATATCGACAA TACGGCGATT
      151  GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
      201  AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
      251  GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAG CGGCAAATGT
      301  ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGCGGCAG AAAACGGCGT
20     351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
      401  GCAAACCTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
      451  GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
      501  AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

25     1  MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI
      51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
      101  METDDKDSPA GWAENGVCHT LFAKLVGNI A EDGGKLT DYL VSHAALQPYQ
      151  AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

30     1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCGC
      51  CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA
      101  GCGCGCCGAA ACCGGTTTTT AAAGTCAAAT ATATCGACAA TACGGCGATT
      151  GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
35     201  AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
      251  GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAG CGGCAAATGT
      301  ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGCGGCAG AAAACGGCGT
      351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
      401  GCAAACCTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
40     451  GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
      501  AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

45     1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
      51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
      101  METDDKDSPA GWAENGVCHT LFAKLVGNI A EDGGKLT DYL VSHAALQPYQ
      151  AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

```

5      orf108.pep  MLNTFFAVLGGCLLXLPCGKSVENTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE  60
      orf108ng    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAAQSAPKPVFKVKYIDNTAIAGLALGQSSE  60

10     orf108.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

15     orf108.pep  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181
      orf108ng    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181

```

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

```

20     orf108-1.pep MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE  60
      orf108ng-1   MLKIPFAVLGGCLLLAACGKSENTAEQPQNAAQSAPKPVFKVKYIDNTAIAGLALGQSSE  60

      orf108-1.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng-1   GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

25     orf108-1.pep  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181
      orf108ng-1   LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181

```

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

```

30      1  ATGCTGAAaA  tacctTTTGC  CGTGTtgggc  ggCtgcctGC  TGCTTGCCGC
      51  CTGCGGCAAA  TCCGAAATa  cggcggaACA  GCCGCAAAAT  gcggCACAAA
     101  GCGCGCCGAA  ACCGGTTTT  AAAGTCAAAT  ACATCGACAA  TACGGCGATT
     151  GCCGGTTTGG  CTTTGGGACA  AAGTAGCGAA  GGCAAAACCA  acgacgGCAA
     201  AAAACAAATC  AGTTATccgA  TTAAAGGCTT  GCCGGAACAA  Aacgcgctcc
     251  gGCTGACCGG  AAAGCATCCC  AACGACTTGG  AagccgtcgT  CGGCAAATGT
     35  301  ATGGAAACCG  ACGGAAAGGA  CGCGCCTTCG  GGCTGGGCGG  AAAACGGCGT
      351  GTGCCATACC  TTGTTTGCCA  AACTGGTGCG  CAATATCGCC  GAAGACGGCG
     401  GCAAACCTGAC  TGATTACCTG  ATTTTCGATT  CCGCCCTGCA  ACCCTATCAG
     451  GCAGGCAAAA  GCGGCTATGC  CGCGGTGCAG  AACGGACGCT  ATGTGCTGGA
     501  AATCGACAGC  GaggggGGCGT  TTTATttccg  ccgccgccat  tattgA

```

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

      1  MLKIPFAVLG  GCLLLAACGK  SENTAEQPQN  AAQSAPKPVF  VKYIDNTAI
     51  AGLALGQSSE  GKTNDGKKQI  SYPIKGLPEQ  NAVRLTGKHP  NDLEAVVGKC
    101  METDGKDAPS  GWAENGVCHT  LFAKLVGNI  AEDGGKLTDL  ISHSALQPYQ
    151  AGKSGYAAVQ  NGRYVLEIDS  EGAFYFRRRH  Y*

```

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

-262-

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGT TTGGTTG CGATGATTGC
 51 CGGATTTATC GATGcgatTg cGggCGGGGG TGGTTTGATT ACGCTGCCCCG
 101 CACTCTTGTT GGCAGGTATT CCTCCCCTGT CGGCAATTGC CACCAACAAG
 151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
 5 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
 201 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
 251 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
 301 CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
 401 TTTTCTGTT cGGGCTGACG GTCGC .ACCG CTTTGGGTT TTTACGACGG
 10 TGTGTTGCGA CCGGGTGTG GCTCGTTTTT TCTGATTGCC TTTATTGTTT
 451 TGCTCGGCTG CAAGCTGTG AACGCGATGT CTTACACCAA ATTGGCGAAC
 501 GTTGCTGCA ATCTTGGTTC GCTATCGGTA TTCCTGCTGC ACGGTTTCGAT
 551 TATTTTCCCG ATTGCGGCAA CGaTGGCGGT CGGTGCGTTT GTCGGTGC GA
 601 ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA
 651

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VXTAFGFLRR
 151 CVRTGCRLVF SDCLYCFARL QAVRDVLHQ IGERCLQSWF AIGIPAARFD
 20 YFPDCGNDGG RCVCRCEERC EICRTLRFEA D*

Further work revealed the following DNA sequence <SEQ ID 433>:

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGT TTGGTTG CGATGATTGC
 51 CGGATTTATC GATGCGATTG CCGGCGGGGG TGGTTTGATT ACGCTGCCCCG
 101 CACTCTTGTT GGCAGGTATT CCTCCCCTGT CGGCAATTGC CACCAACAAG
 25 151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
 251 TAGGCGGCGT GGCCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
 301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
 30 TTTTCTGTT cGGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT
 401 CTGTTCCGAC CCGGTGTGCG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
 451 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG
 501 TTGCCTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
 35 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
 651 TTTAGGTGCG AGATTTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
 701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
 751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
 40 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFDYDG
 151 VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLLHGS
 201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
 251 RNPLYQMIVS MF*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N. meningitidis*:

50	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
	orf109a	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
		10 20 30 40 50 60
55	orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
	orf109a	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
		70 80 90 100 110 120

```

                    130      140      150      160      170      180
orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVRDVLHQ
             |||||
5  orf109a   KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
                    130      140      150      160      170      180

```

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

```

1  ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
10 51 CCGATTATC  GATGCGATTG CGGGTGGGGG TGGTTTGATT ACGCTGCCTG
101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CCGCAATTGC CACCAACAAG
151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCGGCA GCATCGTTTG
251 CAGGCGGCGT GGTCTGGTCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
15 351 GTTTTCGCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTCTGTT CCGTCTGACG GTTGCAACAC TTTTGGGTTT TTACGACGGT
451 GTGTTTCGAC CGGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACC
551 TTGCCTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
20 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
651 TTTAGGTGCG AGATTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAATGTGT GATAGACGAG
751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 436>:

```

25 1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151 VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
30 251 RNPLYQMIVS MF*

```

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

```

                    10      20      30      40      50      60
orf109a.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
35 orf109-1   MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
                    10      20      30      40      50      60

                    70      80      90      100     110     120
orf109a.pep  TVSFARKGLIDWKKGLPIAAASFAGGVVGALS SVSLVSKDILLAVVPVLLIFVALYFVFSP
40 orf109-1   TVSFARKGLIDWKKGLPIAAASFVGGVAGALS SVSLVSKDILLAVVPVLLIFVALYFVFSP
                    70      80      90      100     110     120

                    130     140     150     160     170     180
orf109a.pep  KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
45 orf109-1   KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
                    130     140     150     160     170     180

                    190     200     210     220     230     240
orf109a.pep  LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
50 orf109-1   LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
                    190     200     210     220     230     240

                    250     260
orf109a.pep  SMAVKLLIDERNPLYQMIVSMFX
55 orf109-1   SMAVKLLIDERNPLYQMIVSMFX
                    250     260
60

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N.gonorrhoeae*:

```

5      orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60
      orf109ng    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60

10     orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120
      orf109ng    TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120

15     orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180
      orf109ng    KLDGSKEGKARMSFFLFGLTVATAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180

      orf109.pep  IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRTLRFEAD  231
      orf109ng    IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRPLRFEAD  231

```

20 An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

```

1      MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51     LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101    LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VATAFGFLRR
151    CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD
25     201     YFPDCGNDGG RCVCRCEFR C EICRPLRFEA D*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

```

1      ATGGAAGATT TATACATAAT ACTCGCTTTG GGTGGTGTG CGATGATCGC
51     CGGATTTATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCTG
101    CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAAG
151    CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
201    AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
251    CAGGCGGCGT GGTGCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301    TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
351    GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401    TTTTCTTATT CGGGCTGACG GTTGCAACGC TTTTGGGTTT TTACGACGGT
451    GTGTTGCGAC CGGGTGTGCG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
501    GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG
551    TTGCTTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
601    ATTTTCCCGA TTGTGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
651    TTTAGGTGCG AGATTTCGCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701    TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
751    AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

```

45     1      MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51     LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101    LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151    VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLLHGS
201    IFPIVATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
251    RNPLYQMIVS MF*

```

50 ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

```

10      20      30      40      50      60
orf109ng-1.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
101     MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
55      10      20      30      40      50      60

100     70      80      90      100     110     120
orf109ng-1.pep TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP

```

-265-

```

      |||
orfl09-1  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
              70      80      90      100      110      120
5
      130      140      150      160      170      180
orfl09ng-1.pep KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
      |||
orfl09-1  KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
              130      140      150      160      170      180
10
      190      200      210      220      230      240
orfl09ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      |||
orfl09-1  LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
              190      200      210      220      230      240
15
      250      260
orfl09ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
      |||
orfl09-1  SMAVKLLIDERNPLYQMIVSMFX
              250      260

```

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|94984|pir||I38164 hypothetical protein 9 - Pseudomonas sp >gi|551929
25 (M62866) ORF9 [Pseudomonas denitrificans] Length = 261
   Score = 175 bits (439), Expect = 3e-43
   Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

Query: 41  PPVSAIATNKLQXXXXXXXXXXXXXXXXXKGLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKDI 100
      PP-  + TNKLQ                      R+G ++ K+ LP+                      D+
30 Sbjct: 43  PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFF 160
      L A++P LLI +ALYF  P + G  + +R++ F+F LT+ PL+GFYDGVFGPG GSFF
35 Sbjct: 103 LKAILPFLIAIALYFGLKPNM-GDVDQHSRVTPFVFTLTTLVPLIGFYDGVFGPGTGSFF 161

Query: 161 LIAFIVLLGCKLLNAMSYTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGA 220
      ++ F+ L G  +L A ++TK N  N+G+ VFL G++++ +  M +G F+GA +G+
40 Sbjct: 162 MLGFVTLAGFGVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLGAQVGS 221

Query: 221 RFAVRFGSKLIKPLLIVISISMVAVKLLIDERNPL 254
      R+A+  G+K+IKPLL+++SI++A++LL D  +PL
Sbjct: 222 RYAMAKGAKIIKPLLIVIVISIALAIRLLADPTHPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

```

50      1  ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAACCAG
      51  CCGCAGACGG ATTATTTGGT CAAATTCGGA TCGTTTTGGG CGAG.ATTTT
      101  TGGTTTTCTG GGACTGTATG ACGTCTATGC TTCGGCATGG TTTGTCGTTA
      151  TCATGATGTT TTTGGTGGTT TCTACCACTT TGTGCCTGAT TCGCAATGTG
      201  CCGCCGTTCT GGCGCGAAAT GAAGTCTTTT CGGGAAAAGG TTAAAGAAAA
55      251  ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGATGTA AAAATTGCGC
      301  CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTCA GGGGAAAACC
      351  ATTAACCGTG AAGACGGGTC GGTCTGATT GCCGCCAAA AAGGCACAAT
      401  GAACAAATGG GGCTATATCT TGCCCATGT TGCTTTGATT GTCATTTGCC
      451  TGGCGGGT GATAGACAGT AACCTGCTGT TGAACTGGG TATGCTGACC
60      501  GGTCCGATTG TTCCGACAA TCAGGCGGTT TATGCCAAGG ATTC.AAGC

```

-266-

551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTTAGGG GCAACGTCAA
601 TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5 1 ..LLGIASVIGT LLQONQPOTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI
 51 MMFLVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP
 101 EVAKRYLEVQ GFQGKTINRE DGSVLIAAKK GTMNKWGYIF AHVALIVICL
 151 GGLIDSNLL KLGMLTGRIF RTIRRFMPRI XKPESXFGCV QSLI*GQRQY
 201 FXRGRVRMWF S*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf88a.pep	MSKSRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPOTDYLKFGSFWA					
15	orf110			LLGIASVIGTLLQONQPOTDYLKFGSFWA			
				10	20	30	
		70	80	90	100	110	120
20	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
	orf110	XIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
		40	50	60	70	80	90
25	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL					
	orf110	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL					
		100	110	120	130	140	150
30	orf88a.pep	GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNI SEGQSADVVF					
	orf110	GGLIDSNLLKLGMLTGRIFRTIRRFMPRI XKPESXFGCVQSLIXGQRQYFXRGRVRMWF					
35		160	170	180	190	200	210
	orf88a.pep	LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT					
	orf110	SX					

40 However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45	orf110.pep	LLGIASVIGTLLQONQPOTDYLKFGSFWA	30
	orf110ng	MSKSRISPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPOTDYLKFGFPFWT	60
50	orf110.pep	XIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	90
	orf110ng	RIFDFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL	150
55	orf110ng	SSLLDVKIAPEVAKRYLEVGRGFQGKTVSREDGSVLIAAKKGTMNKWGYIXAHVALIVICL	180

-267-

```

orf110.pep  GGLIDSNLLKLGLMTGRIFRTIRRFMPRIKXPESXFGCVQSLIXGQRQYFXRGRVRMWF  210
| ||: ||||| ||||| ||||: || |||| |||| :| ||||| ||||| ||: |||||
orf110ng    GRLINXNLLKLGLMAGSIFRNRRVMPRIKXPESIWWGGVQSLIKGQRQYFQRGKVRMWF  240

```

```

5  orf110.pep  S  211
   |
   orf110ng    S  241

```

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

```

10      1  MSKSRISPTL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLOQNQPQTD
      51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
     101  REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVVR GFQGKTVSRE
     151  DGSVLIAAKK GTMKNWGYIX AHVALIVICL GRLINXNLLL KLGMLAGSIF
     201  RNNRRVMPRI SKPESIWGGV QSLIKGQRQY FQRGKVRMWF S*

```

15 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

```

20      1  ATGCCGCTCTG AAACACGCCT GCCGAACTTT ATCCGCGTCT TGATATTGTC
      51  CCTGGGTTTC ATCTTCCTGA ACGCCTGTTC GGAACAAACC GCGCAAACCG
     101  TTACCCTGCA AGGCGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT
     151  TCAATAATC GGGACAAACT CCCCTCACCT GCCGAAATAC AAAACGCGAT
     201  CGATGACGCG CTTAAAGAAG TCAACCGGCA GATGTCCACC TATCAGCCCG
     25  251  ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCCTCCGC
      301  ATTTCAAGCG ACTTCGCACA CGTTACTGCC GAAGCCGTCC GCCTGAACCG
      351  CCTGACACAC GGCGCGCTGG ACGTAACCGT CGGCCCTTG GTCAACCTTT
     401  GGGGATTCGG CCCGACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA
     451  ATCAAACAGG CGGCATCTTA TACGGGCATA GACAAAATCA TTTTGAAACA
     501  AGGCAAAGAT TACGTTCTCT TGAGCAAAAC CCACCCCAAG GCCTATTGTC
     551  ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATAAAGT TGCGGGCGGAA
     601  CTGGAAAAAT ACGGCATTCA AAATTATCTG GTCGAAATCG GCGGCGAGTT
     651  GCACGGCAAA GGCAAAAACG CGCGCGGCGA ACCGTGGCGC ATCGGTATCG
     701  AGCAGCCCAA TATCGTCCAA GCGGCGCAATA CGCAGATTAT CGTCCCGCTG
     751  AACAAACGTT CGCTTGCCAC TTCCGGCGAT TACCGTATTT TCCACGTCGA
     801  TAAAAACGGC AAACGCCTCT CCCATATCAT CAACCCGAAC AACAAACGAC
     851  CCATCAGCCA CAACCTCGCC TCCATCAGCG TGGTCGCAGA CAGTGCATG
     901  ACGGCGGACG GCTTGTCAC AGGATTATTC GTATTGGGCG AAACCGAAGC
     951  CTTAAAGCTG GCAGAGCGCG AAAAAGCTCG TGTTTTCCTG ATTGTCAGGG
    1001  ATAAAGGCGG CTACCGCACC GCCATGTCTT CCGAATTGTA AAAACTGCTC
    1051  CGCTAA

```

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

```

45      1  MPSETRLPNF IRVLIFALGF IFLNACSEQT AQTVTLQGET MGTITYVKYL
      51  SNNRDKLPSP AEIQKRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR
     101  ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFPGDK SVTREPSPEQ
     151  IKQAASYTGI DKIIKQGKD YASLSKTHPK AYLDLSSIAK GFGVDKVGAGE
     201  LEKYGIQNYL VEIGGELH GK NARGEPWR IGIEQPNIVQ GGNTQIIVPL
     251  NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVADSAM
     301  TADGLSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
     351  R*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N. meningitidis*:

5	orf111a.pep	10	20	30	40	50	60
	orf111	10	20	30	40	50	60
10	orf111a.pep	70	80	90	100	110	120
	orf111	70	80	90	100	110	120
15	orf111a.pep	130	140	150	160	170	180
	orf111	130	140	150	160	170	180
20	orf111a.pep	190	200	210	220	230	240
	orf111	190	200	210	220	230	240
25	orf111a.pep	250	260	270	280	290	300
	orf111	250	260	270	280	290	300
30	orf111a.pep	310	320	330	340	350	
	orf111	310	320	330	340	350	

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGTCTG	AAACACGCCT	GCCGAACTTT	ATCCGCACCT	TGATATTTGC
	51	CCTGAGTTTT	ATCTTCCTGA	ACGCCTGTTT	GGAACAAACC	GCGCAAACCG
	101	TTACCTGCA	AGGTGAAACG	ATGGGCACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACNAACT	CCCNTCACCT	GCCGAAATAC	AAAANCGCAT
	201	CGATGACGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCACAC	TATCAGCCCG
45	251	ACTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCTCCGC
	301	ATTTCAAGCG	ACTTCGCACA	CGTTACTGCC	GAAGCCGTCC	ACCTGAACCG
	351	CCTGACACAC	GGCGCGCTGG	ACGTAACCGT	CGGCCCTTG	GTCAACCTTT
	401	GGGGATTTCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
50	451	ATCAACAAG	CAGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTGAAACA
	501	AGGCAAGAT	TACGTTCTCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATNANGT	TGCGGGCGAA
	601	CTGGAAAAAT	ACGGCATTCA	AAATTATCTG	GTCGAAATCG	GCGGNGAGTT
	651	GCACGGCAAA	GNCAAAAACG	CGCGGGCGCA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GGCGGCAATA	CGCAGATTAT	CGTCCGCTG
55	751	AACAACCGTT	CGNTTGCCAC	TTCCGGCGAT	TACCGTATTT	TCCACGTCGA
	801	TAAAAGCGGC	AAACGCCTCT	CCCATATCAT	TAATCCGAAC	AACAAACGAC
	851	CCATCAGCCA	CAACCTCGCC	TCCATCAGCG	TGNTCGCAGA	CAGTGCATG
	901	ACGGCGGACG	GCTTNTCCAC	AGGATTATTC	GTATTGGGCG	AAACCGAAGC
	951	CTTAAAGCTG	GCAGAGCGCG	AAAACTCGC	TGTTTCCTG	ATTGTCAGGG
60	1001	ATAAAGGCGG	CTACCGCACC	GCCATGTCTT	CCGAATTTGA	AAAACTGCTC
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

1	MPSETRLPNF	IRTLIFALS	FI	FLNACSEQT	AQTVTLQGET	MGTTYTVKYL
51	SNNRDXLPSP	AEIQXRIDDA	LKEVNRQMST	YQPDSEISRF	NQHTAGKPLR	

-269-

5
101 ISSDFAHVTA EAVHLNRLTH GALDVTVGPL VNLWGFGDPK SVTREPSPEQ
151 IKQAASYTGI DKIIKQKGD YASLSKTHPK AYLDLSSIAK GFGVDXVAGE
201 LEKYGIQNYL VEIGGELHGK XKNARGEPRW IGIEQPNIVQ GGNTQIIIVPL
251 NNRSXATSGD YRIFHVDKSG KRLSHIINPN NKRPIHNLASISVVSADAM
301 TADGXSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
351 R*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

		10	20	30	40	50	60
	orf111ng	MPSETRLPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSP					
	orf111	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSP					
15		10	20	30	40	50	60
	orf111	AKIQKRIDDALKEVNRQMSYQTDSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH					
20	orf111	AEIQKRIDDALKEVNRQMSYQDPSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH					
		70	80	90	100	110	120
	orf111ng	GALDVTVGPLVNLWGFGDPKSVTREPSPEQIKQAASYTGIDKIILOQKGDYASLSKTHPK					
25	orf111	GALDVTVGPLVNLWGFGDPKSVTREPSPEQIKQAASYTGIDKIILOQKGDYASLSKTHPK					
		130	140	150	160	170	180
	orf111ng	AYLDLSSIAKGFGVDKVAGELEKYGIQNYLVEIGGELHGKGNHAGEPWRIGIEQPNIIQ					
30	orf111	AYLDLSSIAKGFGVDKVAGELEKYGIQNYLVEIGGELHGKGNARGEPRWIGIEQPNIVQ					
		190	200	210	220	230	240
	orf111ng	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPIHNLASISVVSADAM					
35	orf111	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPIHNLASISVVSADAM					
		250	260	270	280	290	300
	orf111ng	TADGLSTGLFVLGETEALRLAEQEKLAVFLIVRDKDGYRTAMSSEFAKLLRX					
40	orf111	TADGLSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					
		310	320	330	340	350	
45	orf111ng	TADGLSTGLFVLGETEALRLAEQEKLAVFLIVRDKDGYRTAMSSEFAKLLRX					
	orf111	TADGLSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

	1	ATGCCGTCG	AAACACGCCT	GCCGAACCTT	ATCCGCGCCT	TGATATTTGC
	51	CCTGGGTTTC	ATCTTCCTGA	ACGCCTGTTC	GGAacaaacC	GCGCAaacccg
50	101	TTACCCCTGCA	AGGCGAAAcg	aTGGGTACGA	CCTATACCGT	CAAAATACCTT
	151	TCAAATAATC	GGGACAAACT	CCCCTCCCCT	GCCAAAATAC	AAAAGCGCAT
	201	TGATGATGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TACGAGACCG
	251	ATTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCCTCCGC
	301	ATTTCAAGCG	ATTTTCGACA	CGTTACCGCC	GAAGCCGTCC	GCCTGAACCG
55	351	CCTGACTCAC	GGCGCACTGG	ACGTAACCGT	CGGCCCTTTG	GTCAACCTTT
	401	GGGGGTTTCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAACAGG	CGGCATCTTA	TACGGGCATA	GACAAAATCA	TTTGTCAACA
	501	AGGCAAAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAA	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATAAAGT	TGCGGGCGAA
	601	CTGGAAAAAT	ACGGCATTCA	AAATTATCTG	GTCGAAAtcg	gcggcgGAGTT
60	651	GCACGGCAAA	GGCAAAAATG	CGCACGGCGA	ACCGTGGCGC	ATCGGTATAG
	701	AGCAACCCAA	TATCATCCAA	GgcgGCAata	CGCAGATTat	cgtecccgctg
	751	aaCaaccgtt	cgctTGCCAC	TTCCGGCGAT	TAccgtaTTT	tccacgtcga
	801	TAAAAAcggc	aaacgccttt	cccacaTCAT	CAATCCCaAC	aacAAACgac
	851	ccATCAGcca	caacctcgcc	tccatcagcg	tggctctcAGA	CAGTGCATG
65	901	ACGGCGGACG	GTTTatCCAC	AGGATTATTT	GTTTTAGGCG	AAACCGAAGC
	951	CTTAAGGCTG	GCAGAACAAAG	AAAAACTCGC	TGTTTTCTTA	ATTGTCCGGG

-270-

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTTC CAAGCTGCTC
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

```

5      1  MPSETRLPNL IRALIFALGF IFLNACSEQT AQTVTLOGET MGTITYTVKYL
      51  SNNRDKLPSP AKIQKRIDDA LKEVNRQMSY YQTDSEISRF NQHTAGKPLR
     101  ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFPGDK SVTREPSPSEQ
     151  IKQAASYTGI DKIILOQGGK YASLSKTHPK AYLDLSSIAK GFGVDKVAGE
     201  LEKYGIQNYL VEIGGELHGK GKNAHGEPWR IGIEQPNIIQ GGNTQIIIVPL
     251  NNRSLATSGD YRIFHVVDKNG KRLSHIINPN NKRPISHNLA SISVVSDSAM
    10  301  TADGLSTGLF VLGETEALRL AEQEKLAFLV IVRDKDGYRT AMSSEFAKLL
     351  R*

```

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

```

15  sp|P44550|YOJL_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4
    hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)
    >gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346
    Score = 353 bits (896), Expect = 9e-97
    Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)

20  Query: 7  LPNLIRALIFALGFIFLNACSEQTAQTVTLOGETMGTITYTVKYL SNNRDKLPSPAKIQKR 66
      + LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +
    Sbjct: 1  MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTTYHVKYLDGGSITATSE-KTHEE 58

    Query: 67  IDDALKEVNRQMSYQTDSEISRFNQHT-AGKPLRISSDFAHVTA EAVRLNRLTHGALDV 125
      I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV
    25  Sbjct: 59  IEAILKDVNAKMSTYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTEGALDV 118

    Query: 126  TVGPLVNLWGFPGDKSVTREPSPSEQIKQAASYTGIDKIILOQGGKDYASLSKTHPKAYLDL 185
      TVGP+VNLWGFPG+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL
    30  Sbjct: 119  TVGPVVNLWGFGEKRPKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178

    Query: 186  SSIAGKFGVDKVAGELEKYGIQNYLVEIGGELHGKGNHGEHPWRIGIEQPNIIQGGNTQ 245
      SSIAGKFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +
    35  Sbjct: 179  SSIAGKFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKWPQIAIEKPTTTGERAVE 238
    Query: 246  IIIVPLNNRSLATSGDYRIFHVVDKNGKRLSHIINPNKRPISHNLA SISVVSDSAMTADGL 305
      ++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL
    Sbjct: 239  AVIGLNNMGMASSGDYRIY-FEENGKRFAHEIDPKTGYPHQHHLASITVLAPTSMTADGL 297

    Query: 306  STGLFVLGETEALRLAEQEKLAFLVIVRDKDGYRTAMSSEFAKL 349
      STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL
    40  Sbjct: 298  STGLFVLGEDKALEVAEKNLAVYLIIRTNDNGFVTKSSSAFKKL 341

```

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 54

45 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

```

50      1  ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCGTCA
      51  AAAATTGTGG CTGCGCTTCA TCGGCGGCGG GTCGCATCAA AATATACGGG
     101  GCGGCGCGGC TGCAGACGGG TGGCGCAAAG GCGTGCAAAT CGGCGGCGAG
     151  GTGTTGTGAC GGCAAAATGA AGGCAGCCKA yTGGCAATCG GCGTGATGGG
    50  201  CGGCAGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG
      251  gCAGTGATTT GTATGTTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG
     301  TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA
     351  ACGTTTCAAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA
     401  CCAAAGGTTG GACGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG
    55  451  GAAGGCATTG TCGGAAAAGG CAATAATGTG CGGTTTTACC TACAACCGCA
      501  GgCGCAGTTT ACCTACTTGG GCGTAAACGG CGGCTTTACC GACAGCGAGG
     551  GGACGGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC
     601  AtTCGGGCAA AAACCGTTT TGCTTTGCGT AACGGTGTCA ATCTTGAGCC
     651  TTTTGCCGCT TTTAATGttT TGCACAGGTC AAAATCTTTC GGCGTGGA
    60  701  TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG

```

```

1      ..PCRRQGDDVY  AAHASRQKLW  LRFIGGRSHQ  NIRGGAADG  WRKGVQIGGE
51     VFVRQNEGSX    LGVGMGGRA  GQHASVNGKE  GAAGSDLYG  GGGVYAAWHQ
5      101    LRDKQTGAYL  DAWLQYQRFK  HRINDENRAE  RYKTKGWTA  VEGGYNALVA
151    EGVGKGNV      RFYLQPAQF  TYLGVNGGET  DSEGTAVGLL  GSGQWQSRAG
201    IRAKTRFALR    NGVNLQPF  FNVLHRSKSF  GVEMDGEKQT  LAGRTALEGR
251    FGIEAGWKGH     GSA..

```

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

	Orf35	5	QGGDDVYAAHASRQKLWLRFIGGRSHQNIRGGAA-ADGWRKG V QIGGEVFVRQNEGSXLAI	63
			+ D++ R+ LWLR I G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I	
15	virg-h	396	KNSDIFDRRTLPRKGLWLRVIDGHSNQWVQGKTAPVEGYRKG V QLGGEVFTWQNESNQLSI	455
	Orf35	64	GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDKQTGAYLDGWLQYQRFKH	121
			G+MGG+A Q ++ + ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H	
	virg-h	456	GLMGQAQERSTFHNPDNDNLTTGNVKGFAGVYATWHQLQDKQTGAYADSWMQYQFRH	515
20	Orf35	122	RINDENRAERYKTKGTASVEGGYNALVAEGIVGKGNNVRYFLYPQAQFTYLGVN GGFTD	181
			RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLPQPAQ TYLG VNG F+D	
	virg-h	516	RINTEDGTERFTSKGITASIEAGYNALLAEHFTHKKGNSLRVYLPQAQLTYLG VNGKFSD	575
25	Orf35	182	SEGTAVGLLLSGSQWSRAGIRAKTRFALRNGVNLPFAAFNVLHRSSKSFVEMDGEKQTL	241
			SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVE MDGE++ +	
	virg-h	576	SENAHVNLLGSRQLQTRVG VQAKAQFSLYKNIAIEPF AA VNALYHNKPFGVEMDGERRVI	635
	Orf35	242	AGRTALEGRFGIEAGWK GHMS	262
			+TA+E + G+ K H++	
30	virg-h	636	NNKTAIESQLGVAVKIKSHLT	656

ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N. meningitidis*.

BNSDOCID: <WO 9924578A2 | >

orf35a LQPF AAFNV LHRSKSFGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSARIGYGKRTDGD
550 560 570 580 590 600

5 orf35a KEAALS LKWLFX
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

```

1  ATGTTTCAGAG CTCAGCTTGG TTCAAATACT CGTTCTACCA AAATCGGCGA
51  CGATGCCGAT TTTTCATTTT CAGACAAGCC GAAACCGGGC ACTTCCCATT
101 ATTTTTCCAG CCGTAAAACC GATCAAAATT CATCCGAATA TGGGTATGAC
151 GAAATCAATA TCCAAGGTAA AAATAACAAT AGCGGCATAC TCGCCGTCGA
201 TAATATGCCC GTTGTTAAGA AATATATTAC AGATACTTAC GGGGATAATT
251 TAAAGGATGC GGTTAAGAAG CAATTACAGG ATTTATACAA AACAAGACCC
301 GAAGCTTGGG AAGAAAATAA AAAACGGACT GAGGAGGCGT ATATAGAACA
351 GCTTGGACCA AAATTTAGTA TACTCAAACA GAAAACCCCC GATTTAATTA
15 401 ATAAATTGGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA
501 CAAATCCAC GTCGCCGGAC AGGTGTTGGA ACTGACCAAG ATGACGCTGA
551 AAGATTCCCT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA
601 ACTTCCGATA ATGCCCGCAT CCGCCTGAAC ACGAAAGATG AAAAAGTAC
20 651 CGTCCATAAA CGGTATCAGG GCGGTGCGGA TTTCTGTTC GGCTACGACG
701 TGCGGGAGTC GGACAACCC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA
751 CAATCCGGCG TGGTTTTGGA ACGCCGGCCG GAAAATCTGA AAACGCTCGA
801 CGGGCGCAAA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTGCGT
851 TTAAACAAAA TTACCGGCAG GGAATGTACG AATTATTGCT CAAGCAATGC
25 901 GAAGGCGGAT TTTGCTTGGG CGTGCAGCGT TTGGCTATCC CCGAGGCGGA
951 AGCGGTTTTA TATGCCCAAC AGGCTTATGC GGCAATACT TTGTTGCGGC
1001 TGCGTGCCGC CGACAGGGC GACGACGTGT ATGCCGCCGA TCCGTCCCGT
1051 CAAAAATTGT GGCTGCGCTT CATCGGCGGC CGGTGCGATC AAAATATACG
1101 GGGCGGCGCG GCTGCGGACG GGCGGCGCAA AGGCGTGCAA ATCGGCGGCG
30 1151 AGGTGTTTGT ACGGCAAAAT GAAGGCAGCC GGCTGGCAAT CCGCGTGATG
1201 GCGGCAGGG CTGGCCAGCA CGCATCAGTC AACGGCAAAG GCGGTGCGGC
1251 AGGCAGTTAT TTGCATGGTT ATGGCGGGGG TGTTTATGCT GCGTGGCATC
1301 AGTTGCGCGA TAAACAAACG GGTGCGTATT TGGACGCTG GTTGCAATAC
1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGCGG AACGCTACAA
35 1401 AACCAAAGT TGGACGGCTT CTGTGCAAGG CGGCTACAAC GCGCTTGTGG
1451 CGGAAGGCGT TGTCGGAAAA GGCAATAATG TGCGGTTTTA CCTGCAACCG
1501 CAGGCGCAGT TTACCTACTT GGGCGTAAAC GGGCGGTTTA CCGACAGCGA
1551 GGGGACGGCG GTCGGACTGC TCGGCAGCGG TCAAGTGCAA AGCCGCGCCG
1601 GCATTGCGGC AAAAACCCGT TTTGCTTTGC GTAACGGTGT CAATCTTCAG
40 1651 CCTTTTGCCG CTTTTAATGT TTTGCACAGG TCAAAATCTT TCGGCGTGGA
1701 AATGGACGGC GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGGC
1751 GGTTGCGCAT TGAAGCCGGT TGGAAAGGCC ATATGTCCGC ACGCATCGGA
1801 TACGGCAAAA GGACGGACGG CGACAAAGAA GCCGCATTGT CGCTCAAATG
1851 GCTGTTTTGA

```

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

```

1  MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT DQNSSEYGYD
51  EINIQGKNYN SGILAVDNMP VVKYITDITY GDNLKDAVKK QLQDLYKTRP
101  EAWEEKKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ
151  TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RHRSDIHMLE
50 201  TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEEKVSG
251  QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLLKQC
301  EGGFCLGVQR LAIPEAEAVL YAQQAYAANT LFGLRAADRG DDVYAADPSR
351  QKLWLRFIGG RSHQNIIRGA AADGRRKGVQ IGGEVFRQON EGSRLAIGVM
401  GGRAGQHASV NGKGAAGSY LHGYGGGVYA AWHQLRDKQT GAYLDGWLQY
55 451  QRFKHRINDE NRAERYKTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLPQ
501  QAQFTYLGVN GGFTDSEGT VGLLGSQWQ SRAGIRAKTR FALRNGVNLQ
551  PFAAFNVLHR SKSFGVEMDG EKQTLAGRTA LEGRFGIEAG WKGHMSARIG
601  YGKRTDGDKE AALSLKWLFX*

```

Homology with a predicted ORF from *N. gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N. gonorrhoeae*:

```

orf35.pep      PCRRQGDVYAAHASRQKLWLRFIGGRSHQNIIRG      34
               :::::  |::| |::| |::|
orf35ngh      FTKVQERDDIAIYAQQAQAANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG      370

```

[illegible]

A partial ORF35n_{gh} nucleotide sequence <SEQ ID 455> is predicted to encode a protein having partial amino acid sequence <SEQ ID 456>:

20	1	..KKLRDRNSEY	WKEETYHIKS	NGRTYPNIPA	LFPKHPDFPF	ENINNSKKIS
	51	FYDKEYTEDY	LVGFARGFGV	EKRNGEEKEP	LRQYFKDCVN	TENSNDNDCK
	101	ISSFGNYGP	LIKSDIFALA	SQIKNSHINS	EILSVGNIE	WLRPTLNKLT
	151	GWQEHLYAGL	DPFHYIEVTD	NSHVIGQITD	LGAEALNTSL	WKPRWNSNID
	201	YLITKNAEIR	FNTKNESLLV	KEDYAGGARF	RFAYDLKDKV	PEIPVLTFEK
25	251	NITGTSDIIF	EGKALDNLKH	LDGHQIVKVN	DTADKDAFRL	SSKYRKGITY
	301	LSLQQRPEGF	FTKVQERDDI	AIYAQQAQAA	NTLFALRLND	KNSDIFDRTL
	351	PRKGLWLRLV	DGHSNQWVQG	KTAPVEGYRK	GVQLGGEVFT	WQNESNQLSI
	401	GLMGGQAEQR	STFRNPDTDN	LTTGNVKGFG	AGVYATWHQL	QDKQGTAYVD
	451	SWMQYQRFRH	RINTEYATER	FTSKGITASI	EAGYNALLAE	HFTKKGNSLR
30	501	VYLQPQAQLT	YLGVNGKFSD	SENAQVNLLG	SRQLQSRVGV	QAKAQFAFTN
	551	GVTFPQFVAV	NSIYQQKPFQ	VEIDGDRRVI	NNKTVIETQL	GVAAKIKSHL
	601	TLOASENROT	SKHHHAKOGA	LNLQWTF*		

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 55

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

40

```

      1  . . GCGGAATATG  TTCAGTTCTC  TATAGATTTG  TTCAGTGTGG  GTAAATCGGG
     51  GGGCGGTATA  CCTAAGGCTA  AGCCTGTGTT  TGATGCGAAA  CCGAGATGGG
    101  AGGTTGATAG  GAAGCTTAAT  AAATTGACAA  CTCGTGAGCA  GGTGGAGAAA
    151  AATGTTCAGG  AAACGAGAAG  AAGGAGTCAG  AGTAGTCAGT  TTAAGCCCCA
    201  TGCGCAACGA  GAATGGGAAA  ATAAAAACAG  GTTAGATTTT  AATCATTTTA
    251  TAGGTGTTGA  TATCAATAAA  AAAGGCACAG  TAACAGGAGG  GCATAGTCTA
    301  ACCCGTGGTG  ATGTACGGGT  GATACAACAA  ACCTCGGCAC  CTGATAAACA
    351  TGGGGT. TTA  TCAAGCGACA  GTGGAAATTN  A
```

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

45 1 ..AEYVQFSIDL FSVGKSGGGI PKAKPVFDAK PRWEVDRKLN KLTTRQVEK
 51 NVQETRRRSQ SSQFKAHAQR EWENKTGLDF NHFIGGDINK KGTVTGGHSL
 101 TRGDVRVIOO TSAPDKHGXL SSDSGNX

Further work revealed further partial nucleotide sequence <SEO ID 459>:

50	1	..GCAGTGTGCC	TnCCGATGCA	TGCACACGCC	TCAnATTGTGG	CAAACGATTC
	51	TTTTATCCGG	CAGGTTCTCG	ACCGTCAGCA	TTTCGAACCC	GACGGGAAAT
	101	ACCACCTATT	CGGCAGCAGG	GGGGAACCTTG	CCGAGCGCCA	GTCTCATATC
	151	GGATTGGGAA	AAATACAAAG	CCATCAGTTG	GGCAACCTGA	TGATTCAACA
55	201	GGCGGCCATT	AAAGGAAATA	TCGGCTACAT	TGTCGCGTTT	TCCGATCACG
	251	GGCACGAAGT	CCATTCCCCs	TTCGACAACC	ATGCCTCACA	TTCCGATTCT
	301	GATGAAGCCG	GTAGTCCCGT	TGACGGATTT	AGCCTTTACC	GCATCCATTG
	351	GGACGGATAC	GAACACCATC	CCGCCGACGG	CTATGACGGG	CCACAGGGCG
	401	CGCGCTATCC	CGCTCCCCAA	GGCGCGAGGG	ATATATACAG	TTACGACATA

-274-

5
 451 AAAGGCGTTG CCCAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC
 501 CGGACAACGG CTTGCCGACC GTTCCACAA TGCCGGTAGT ATGCTGACGC
 551 AAGGAGTAGG CGACGGATTC AAACGCGCCA CCCGATACAG CCCCAGAGCTG
 601 GACAGATCGG GCAATGCCGC CGAAGCCTTC AACGGCACTG CAGATATCGT
 651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

10
 1 ..AVCLPMHAHA SXLANDSFIR QVLDRQHFEP DGKYHLFGSR GELAEROSHI
 51 GLGKIQSHQL GNLMIQQAII KGNIGYIVRF SDHGHEVHSP FDNHASHSDS
 101 DEAGSPVDGF SLYRIHWDGY EHHPADGYDG PQGGGYPAFK GARDIYSYDI
 151 KGVAQNIRLN LTDNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSPEL
 201 DRSGNAAEAF NGTADIVKNI IGAAGEI

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

20
 orf46.pep AEYVQFSIDLFSVGKSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 45
 orf46ng PKTGVPFDGKGFPNFEKHVKYDTKLDIQELSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 217
 orf46.pep EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV 105
 orf46ng EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGAVTGGHSLTRGDV 277
 25
 orf46.pep RVIQOTSAPDKHGXLSDDSGN 126
 orf46ng RVIQOTSAPDKHGVLSDDSGN 298

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

30
 1 ..RRLKHCCCHAR LGSFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPKQC
 51 RTRHRSRQQY LYGSHPHQRD WSCPGKIQLG RHGHTSCRAV ADXRDRICER
 101 EIRRQRQXCR CRLGKIPSL S IPKYPLKLEQ RYKGENITSS TVPPSNGKNV
 151 KLADQRHPKT GVPFDGKGFP NFEKHVKYDT KLDIQELSGG GIPKAKPVFD
 201 AKPRWEVDRK LNKLTTRQV EKNVQETRRR SQSSQFKAHA QREWENKTGL
 251 DFNHFIGGDI NKKGAVTGGH SLTRGDV RVI QOTSAPDKHG VLSSDDSGN*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

40
 1 TTGGGCATTT CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG
 51 CCTGCCGATG CATGCACACG CCTCAGATTT GGcaAACGAT CCCTTTATCC
 101 GgCaggttcT CGaccGTCAG CATTTCGaac ccgacggGaa ATACCaCCTA
 151 TtcggCaGCA GGGGGGAGCT TgccnagcGC aacggccATa tcggattggG
 201 aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcggccg
 251 ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa
 301 ttccattcgc ccttcGAcAa ccaTGCCTCA CATTCCGATT CTGACGAAGC
 351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT
 401 ACGAACACCA TCCCCCGGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT
 451 CCCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAAGGCGT
 501 TGCCCAAAAT ATCCGCTCA ACCTGACCGA CAACCGCAGC ACCGGACAAC
 551 GGCTTGCCGA CCGTTTCCAC AATGCCGGCG CTATGCTGAC GCAAGGAGTA
 601 GCGGACGGAT TCAACGCGC CACCCGATAC AGCCCGGAGC TGGACAGATC
 651 GGGCAATGcC gccGAAGCCT TCAACGCGAC TGcAGATATC GTCAAAAACA
 50
 701 TCATCGGCGC GGCAGGAGAA ATTGTCGGCG CAGGCGATGC CGTGcagGGT
 751 ATAAGCGAAG GCTCAACAT TGCTGTCATG CACGGCTTG GTCTGCTTTC
 801 CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC
 851 TCAAAGACTA TGCCGACGA GCCATCCGCG ATTGGGCAGT CCAAAACCCC
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA TGGGACCCAT
 55
 951 CCCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA
 1001 TCACGGCACA TCCTGTCAAG CGGTGCGAGA TGGGCGGAT CGCATTTGCCG
 1051 AAAGGGAAAT CCGCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA
 1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC

-275-

1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC
 1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
 1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT
 1351 GTGTTTGTAT CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
 1401 GACAACCTCG GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA
 1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAATAATAA
 1501 ACAGGGTTAG ATTTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG
 1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC
 1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGAA
 1651 ATTAAAAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA
 1701 AGTGATGACC AAGCACACCA TGTTCCTCAA AGATTGGGAT GAGGCTAGAA
 1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
 1801 AATAAATGGC AGGGTACAAG TAAATCGGGT ATTAAATAG AAGGATTATC
 1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKYHL
 51 FGSRGELAXR NGHIGLGNIO SHQLGHLMIQ QAAVEGNIGY IVRFSDHGHK
 101 FHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD GYDGPQGGGY
 151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLADRFH NAGAMLTQGV
 201 GDGFKRATRY SPELDRSGNA AEAFTGTADI VKNIIIGAAGE IVGAGDAVQG
 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQIEAVS NIFMAAPIK GIGAVRGKYG LGGITAHVPK RSQMGAIALP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
 401 KNVKLADQRH PKTGVPFDGK GFPNFEKHVK YDTKLDIQEL SGGGIPKAKP
 451 VFDAPRWEV DRKLNKLTR EQVEKNVQET RRRSQSSQFK AHAQREWENK
 501 TGLDFNHFIG GDINKKGTVT GGHSLTRGDV RVIQOTSAPD KHGVYQATVE
 551 IKKPDGSWEV KTKKGGKVM THTMFPKDW EARRAEVTS AWESRIMLKD
 601 NKWQTSKSG IKIEGFTEPN RTAYPIYE*

ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

		10	20	30	40
orf46-1.pep		AVCLPMHAHASXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER			
orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR				
		10	20	30	40
		50	60	70	80
orf46-1.pep	QSHIGLGKIQSHQLGNLMIQQAIAIKGNIGYIVRFSDHGHGHEVHSPFDNHASHSDSDEAGSP				
orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAIAVEGNIGYIVRFSDHGHKFHSPFDNHASHSDSDEAGSP				
		70	80	90	100
		110	120	130	140
orf46-1.pep	VDGFSLYRIHWDGYEHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRNLTDNRS				
orf46ng-1	VDGFSLYRIHWDGYEHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRNLTDNRS				
		130	140	150	160
		170	180	190	200
orf46-1.pep	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSGNAEAFTGTADIVKNIIIGAAGE				
orf46ng-1	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSGNAEAFTGTADIVKNIIIGAAGE				
		190	200	210	220
		230	240	250	260
orf46-1.pep	I				
orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAIRDWAVQNP				
		250	260	270	280
		290	300		

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of *N. meningitidis*:

10 20 30 40 50 60

-276-

	orf46a.pep	LGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR
5		10 20 30 40 50 60
	orf46a.pep	70 80 90 100 110 120
	orf46ng-1	SGHIGLGNIQSHQLGNLFIQQAIAKGNIGYIVRFS DHGHEVHSPFDNHASHSDSDEAGSP
10		70 80 90 100 110 120
	orf46a.pep	130 140 150 160 170 180
	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGGYPA PKGARDIYSYDIKGVAQNIRLNLTDNRS
15		130 140 150 160 170 180
	orf46a.pep	190 200 210 220 230 240
	orf46ng-1	TGQRLVDRFHNTGSMLTQGVGDGFKRATRYSP ELDRSGNAAEFNGTADIVKNIIGAAGE
20		190 200 210 220 230 240
	orf46a.pep	250 260 270 280 290 300
	orf46ng-1	IVGAGDAVQGISSEGSNIAMVHGLGLLSTENK MARINDLADMAQLKDYAAAAIRDWAVQNP
25		250 260 270 280 290 300
	orf46a.pep	310 320 330 340 350 360
	orf46ng-1	NAAQGIEAVSNIFTAVIPVKIGAVRGKYGLGG ITAHPVKRSQMGEIALPKGKSAVSDNF
30		310 320 330 340 350 360
	orf46a.pep	370 380 390 400 410 420
	orf46ng-1	ADAAYAKYPSPYHSRNRIRSNEQRYGKENITS STVPPSNGKNVKLANRHPKTKVPFDGK
35		370 380 390 400 410 420
	orf46a.pep	430 440 450 460 470
	orf46ng-1	GFPNFEKDVKYDTRINTAVPQVN---PIDEPV FN--PKGSVGSASHWSITARIQYAKLP
40		430 440 450 460 470
	orf46a.pep	480 490 500 510 520 530
	orf46ng-1	RQGRIRYIPPKNYSPSAPLPKGPNNGYLDKFG NEWTKGPSRTKGQEFWDVQLSKTGREQ
45		480 490 500 510 520 530
	orf46a.pep	QETRRRSQSSQFKAHAQREWENKTGLDFNHFI GGDINKKGTVTGGHSLTRGDVRVIQOTS
50		480 490 500 510 520 530

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATTT	CCCGCAAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATTT	GGCAAACGAT	TCTTTTATCC
	101	GGCAGGTTCT	CGACCGTCAG	CATTTCTGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGGGGAAC	TGCCGAGCGC	AGCGGTCATA	TCCGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCCAG	CAGGCGGCCA
60	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGGCACGAA
	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
	451	CCCGCTCCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAAGGCGT
	501	TGCCCAAAAT	ATCCGCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
65	551	GGCTTGTCGA	CCGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
	701	TCATCGGCGC	GGCAGGAGAA	ATTGTGCGCG	CAGGCGATGC	CGTGCAGGGT
	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTTATG	CACGGCTTGG	GTCGTCTTTC
70	801	CACCGAAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC

-277-

5 851 TCAAAGACTA TGCCGCAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA CGGCAGTCAT
 951 CCCCCTCAA GGGATTGGAG CTGTTCCGGG AAAATACGGC TTGGGCGGCA
 10 1001 TCACGGCACA TCCTGTCAAG CGGTCGCAGA TGGGCGAGAT CGCATTGCCG
 1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA
 1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC
 1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGA
 1201 AAGAATGTGA AACTGGCACA CAAACGCCAC CCGAAGACCA AAGTGCCGTT
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAAAA AGACGTAAAA TACGATACGA
 1301 GAATTAATAC CGCTGTACCA CAAGTGAATC CTATAGATGA ACCCGTCTTT
 1351 AATCCTAAAG GTTCTGTCGG ATCGGCTCAT TCTTGGTCTA TAACTGCCAG
 1401 AATTCAATAC GCAAAATTAC CAAGGCAAGG TAGAATCAGA TATATCCCAC
 1451 CTAAAAATTA CTCTCCTTCA GCACCGCTAC CAAAAGGACC TAATAATGGA
 1501 TATTTGGATA AATTTGGTAA TGAATGGACT AAAGGTCCAT CAAGAACTAA
 15 1551 AGGTCAAGAA TTTGAATGGG ATGTTCAATT GTCTAAAAACA GGAAGAGAGC
 1601 AACTTGGATG GGCTAGTAGG GATGGTAAGC ATTTAAATAT ATCAATTGAT
 1651 GGAAAGATTA CACACAAATG A

This corresponds to the amino acid sequence <SEQ ID 466>:

20 1 LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRO HFEPDGKYHL
 51 FGSRGELAER SGHIGLGNIQ SHQLGNLFIQ QAAIKGNIGY IVRFSHDHGE
 101 VHSFPDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
 151 PAPKGARDIY SYDIKGVAQN IRLNLTNRS TGQRLVDRFH NTGSMLTQGV
 201 GDGFKRATRY SPELDRSGNA AEAFTGTADI VKNIIIGAGE IVGAGDAVQG
 25 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQGIEAVS NIFTAVIPVK GIGAVRGKYG LGGITAHVPK RSQMGEIALP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
 401 KNVKLANKRH PKTKVPFDGK GFPNFEKDVK YDTRINTAVP QVNPIDEPVF
 451 NPKGSVGSAA SWSITARIQY AKLPRQGRIR YIPPKNYSPS APLPKGPNNG
 501 YLDKFGNEWT KGPSRTKGQE FEWDVQLSKT GREQLGWASR DGKHLNISID
 30 551 GKITHK*

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

40 1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCT
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCCTTTCGT
 201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GCGCGTTTTG TTTGACGGGC
 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301 AACCTCGTCC CTTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
 351 CGGGCTG...

45 This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWLAL LTATARPIVN
 51 LDYLPALLI ALPWRFVKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
 101 NLVPFILTAP APYQIMTGL...

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

50 1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCT
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCCTTTCGT
 201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GCGCGTTTTG TTTGACGGGC
 55 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301 AACCTCGTCC CTTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
 351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG

-278-

5
10
15
20

```

401 CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTTCAC CGGCCATTG AGTTACTACG ACCGGGGTGC
501 GATGGCCAAT ATCTTCGGCG CAAACAACCT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGCGG
851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC
901 TGCCTCCCA ACCGTTTGA ACAAGAAGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGGC AGTTCGCTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAGTCTGC
1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTCGGCG AAGTGTGCGG
1101 ATTTTTCAAA AAACACGACA AGGGACTGTT TTAAGTATG ACGCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCGCG CGAAACCGAC CTCTGCCGCA ATTTACGCTT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
1301 TGAAAGGCAT GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTGGC
1351 AACCTCAATG AAACCTTCCG CTACCTCAAA CAGGGGCACG TCGCTGGCT
1401 GAACTCAAA ATCAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

25
30

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLLILLA PNAVFWLAL LTATARPIVN
51 LDYLPALLI ALPWRFKIA GVLAFLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QORAATHLNE PKSQILFIV AESWGLPANP ELQATFAKL
251 LAQKDRFSVW ESGSFPIGA TVEGEMREL CAYGGLRGFAL RRAPDEKFA
301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQIKT AENLIGKTC
351 AIFGGVCDSE LFGEVSAFFK KHKGLFYWM TLTSHADYPE SDIFNHLKC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVPG
451 NLNETFRYLK QGHVAWLNEK IK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N. meningitidis*:

40
45
50

```

              10      20      30      40      50      60
orf48.pep    MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWLALLTATARPIVNLDYLPALLI
              ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40 orf48a     MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVFWLALLTATARPIVNLXYLPAALLI
              10      20      30      40      50      60

              70      80      90      100     110     119
orf48.pep    ALPWRFKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL
              ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 orf48a     ALPWRXVKIXGLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFIXTAPALYQIMTGLL
              70      80      90      100     110     120

50 orf48a     LLYMLAMPFVLQKAAKTDFRHIAACA AVVVAAGYFTGHL SXDRGRMANIFGANNFYA
              130     140     150     160     170     180

```

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

55
60

```

1 ATGAATATTC ACACCCTGCT CTCCAACAA TGGACGCTGC CGCCATTCTCT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGANTACC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTNTCGT
201 CAAAATTGNC GCGGTATTGG CGTNTTGGCT GGCGGTTTTG TTTGACGGGC
251 TGATGATGGT GATCCAACCT TCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCNT GACCGCCCCC GCCCTTTATC AGATAATGAC
351 CGGGCTGTTA CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
60 401 CCGCCGCCAA AACCGACTTC CGACACATTG CCGCTGTGC CGCCGTTGTG
451 GTGGCAGCCG GCTATTTTAC CGGCCATTG AGTTANTACG ACCGGGGGCG

```

-279-

501 GATGGCCAAT ATCTTCGGCG CAAACAACCTT CTATTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGANCGTTT TTCGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGATCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC
901 TGCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTTCGGC AAGTGTCCGC
1101 ANTTTTCAAA AAACACGACA AGGGACTGTT TTAAGTGATG ACGCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCNGACATT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC NTCTGCCGCA ATTTCCAGCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCAA CGCCCGGAAA
1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
1351 AACCTCAATG AAACCTTCCG CTACCTCAA CAGGGGCACG TCGNCTGGCT
1401 GAACTTCAAA ATCAAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLLILLX PNAVFWVLAL LTATARPIVN
51 LXYPALLI ALPWRXVKIX GVLAXWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFIXTAP ALYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAACAAVV
151 VAAGYFTGHL SXYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQATFAKL
251 LAQKXRFVSW ESGSFPIGA TIEGEMREL CAYGGLRGFAL RRAPDEKFA
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKKT
351 AIFGGVCDSE LFGEVSAXFK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC
401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVPG
451 NLNETFRYLK QGHVXWLNFK IK*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
35	orf48a.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNNAVFWVLALLTATARPIVNLXYPALLI					
	orf48-1	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLDPYLPALLI					
		10	20	30	40	50	60
40	orf48a.pep	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFIXTAPALYQIMTGLL					
	orf48-1	ALPWRFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL					
		70	80	90	100	110	120
45	orf48a.pep	LLYMLAMPFVLQKAAAKTDFRHIAACAAVVVAAGYFTGHLSXYDRGRMANIFGANNFYA					
	orf48-1	LLYMLAMPFVLQKAAAKTDFRHIAVCAAVVAAAGYFTGHLSYYDRGRMANIFGANNFYA					
		130	140	150	160	170	180
50	orf48a.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
	orf48-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
		190	200	210	220	230	240
55	orf48a.pep	ELQATFAKLLAQKXRFVSWESGSFPFIGATIEGEMRELCAYGGLRGFALRRAPDEKFA					
	orf48-1	ELQATFAKLLAQKDRFSVWESGSFPFIGATVEGEMRELCAYGGLRGFALRRAPDEKFA					
60		250	260	270	280	290	300
65	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAENLIGKKTCAIFGGVCDSE					
	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAENLIGKKTCAIFGGVCDSE					
		310	320	330	340	350	360

-280-

		370	380	390	400	410	420
orf48a.pep		LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHLRKCTEYGLPAETDXCRNFSLHTQ					
orf48-1		LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHLRKCTEYGLPAETDLNRNFSLHTQ					
		370	380	390	400	410	420
orf48a.pep		FFDQLADLIQRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX					
orf48-1		FFDQLADLIQRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX					
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*

15 *gonorrhoeae*:

orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVEFWVLALLTATARPVNLDYLPALLI	60
orf48ng	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVEFWVLALLTATARPVNLDYLPALLI	60
orf48.pep	ALPWRFVKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	119
orf48ng	ALPWRFVKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	PNAVEFWVLAL	LTATARPVNL
	51	LDYLPALLI	ALPWRFVKIA	GVLAFLAVL	FDGLMMVIQL	FPFMDLIGAI
	101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LQAAVKTD	RHIAVCAAVV
	151	AAARYFTGPF	ELLRTGGRWQ	YVQHRRLLS	GSRASFRRRQ	KADVLRLRLGN
	201	PYASMGNGG				

30 Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCTCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCGG
	101	TGTTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCCGCCC	GATTGTCAAT
	151	TTGGACTACC	TTCCCGCCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
35	201	CAAAATTGCC	GGCGTATTGG	CGTTTTGGCC	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAACTC	TTCCCTTTTA	TGGACCTCAT	CGGCGCCATC
	301	AACCTCGTCC	CCTTCATCCT	GACCGCCCCC	GCCCCCTATC	AGATAATGAC
	351	CGGCTGTGTT	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAAAAAG
	401	CCGCCGTCAA	AACCGACTTC	CGACACATTG	CCGCTGTGTC	CGCCGTTGTG
40	451	GCGGCAGCCG	GCTATTTCAC	CGGCCATTTC	AGTTACTACG	ACCGGGGGCG
	501	GATGGCCAAT	ATCTTCGGCG	CAACAACCTT	CTATTACGCG	aAAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGgcctG
	601	GTCGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CCGCCACGCG
	651	GCTGAGTGAG	CCGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
45	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGACGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCCTT
	801	CATCGGCGCG	ACGGTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
	851	GTTTGC GCGG	GTTTCGACTG	CGCCGCGCGC	CCGACGAAAA	ATTTGCCCGC
	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
50	951	CGGCGCGGGT	AGTTTCGTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
	1001	GCTTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTTCGGC	AAGTGTCCGC
	1101	ATTTTCAAAA	AAACACGACA	AGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
55	1201	ACCGAATACG	GCCTGCCCGC	CGAAACCGAC	CTCTGCCGCA	ATTTACGCCT
	1251	GCACACCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCCCGCGGC
	1351	AACCTCAATG	AAACCTCCG	CTACCTCAAA	CAGGGACACG	TGCCTGGCT
	1401	GCACTTCAAA	ATCAAATAA			

60 This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

1 MNIHALLSEQ WTLPPFLPKR LLLSLILLA PNAVFWLAL LTATARPIVN
 51 LDYLPALLI ALPWRVFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI
 101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD F RHIAVCAAVV
 151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
 5 201 VDPVFLPLGN QQRAATRLSE PKSQKILFIV AESWGLPGNP ELQATFAKL
 251 LAQKDRFSVW ESGSFPIGA TVEGEMREL CAYGGLRGFAL RRAPDEKFA
 301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQIKT AENLIGKKT
 351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLK
 401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPV
 10 451 NLNETFRYLK QGHVAWLHFK IK*

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

		10	20	30	40	50	60
15	orf48-1.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNL	DYLPALLI				
	orf48ng-1	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNL	DYLPALLI				
		10	20	30	40	50	60
20	orf48-1.pep	ALPWRVFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLV	VPFILTAPAPYQIMTGLL				
	orf48ng-1	ALPWRVFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLV	VPFILTAPAPYQIMTGLL				
		70	80	90	100	110	120
25	orf48-1.pep	LLYMLAMPFVLQKAAAKTDFRHIAVCAAVVAAAGYFTGHLSY	DRGRMANIFGANNFYA				
	orf48ng-1	LLYMLAMPFVLQKAAVKTD	FRHIAVCAAVVAAAGYFTGHLSY	DRGRMANIFGANNFYA			
		130	140	150	160	170	180
30	orf48-1.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLNEPKS	QKILFIVAESWGLPANP				
	orf48ng-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATRLSEPKS	QKILFIVAESWGLPGNP				
		190	200	210	220	230	240
35	orf48-1.pep	ELQATFAKLLAQKDRFSVWESGSFPFIGATVEGEMRELCA	YGGLRGFALRRAPDEKFA				
	orf48ng-1	ELQATFAKLLAQKDRFSVWESGSFPFIGATVEGEMRELCA	YGGLRGFALRRAPDEKFA				
		250	260	270	280	290	300
40	orf48-1.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEI	KTAENLIGKKTCAIFGGVCDSE				
	orf48ng-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQIK	TAENLIGKKTCAIFGGVCDSE				
		310	320	330	340	350	360
45	orf48-1.pep	LFGEVSAFFKKHDKGLFYWM	TLTSHADYPESDIFNHLKCTEYGLPAETDL	CRNFSLHTQ			
	orf48ng-1	LFGEVSAFFKKHDKGLFYWM	TLTSHADYPESDIFNHLKCTEYGLPAETDL	CRNFSLHTQ			
		370	380	390	400	410	420
50	orf48-1.pep	FFDQLADLIR	RPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVAWL	NFKIKX			
	orf48ng-1	FFDQLADLIR	RPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVAWL	NFKIKX			
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

      1 ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCCAAAA TCATCATCGT
      51 TACTTTGAGT ATCGCCACGC TTGCCGCCGC CGGCATCGCT ATGTCGCGCG
5      101 GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
      151 GGT TTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
      201 AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
      251 CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
      301 AGTGC GGTTT TGGCTTTGGT TTTCTTGCA CTGGGCGC.G TAGCGCCGAA
10     351 CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAAATAT AACGGGCAAT
      401 TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

      1 ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMOSD FIEPTPWTLA
      51 GLGFLIALMG WMPAPIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
15     101 SAVLALVFLA LGXVAPNGNG XTVQMAGGKY NQQLINMYA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

      1 ATGTCCGAAC AACATATTTT GACTTGGAAG AGTAAATCA ACGCATTGGG
      51 TCCGGGGATC ATGATGGCTT CGGCGGCGGT CGGCGGTTTC CACCTGATTG
      101 CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
20     151 ATCCTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGCATTA
      201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
      251 GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
      301 AACGCGGGCG CGGTCGCCAT TGTAAACGCC GCCATCGTCA AAATGGCGAT
      351 TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
25     401 CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
      451 TCCAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCGCG
      501 CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
      551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
      601 ATGCCGCGCG CGATTGAAAT TTCCGCCATC AATTCTTTGT GGTAACCGA
30     651 AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTC
      701 ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTT CTTGCACTG
      751 GCGCGCTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
      801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT
      851 GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCAGC
35     901 ACGATTACCG CTATGCGCGG CTATGCCCGT GCCATTGCCG AACCGCTGCG
      951 CCTGCTGCGC GGAAAAGACA AAACGGGCAA CGCCGAATTC TTTGCTTGA
100    1001 ATATTGCGGT GCGGGGCGAG GGTGCGCGG TGATTTTCTG GTTGACGGC
      1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCGCG
      1101 CCCTGTGTTT CGCTGGCTGA ATTACCGTTT GGTTAAAGGT GATGAAAAC
40     1151 ACAAACTCAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTAT
      1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
      1251 ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

      1 MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASQAGA LYGWQIALII
45     51 ILTNLFKYPF FRFSAHYTLD TGKSLIEGYA EKSRYVLWVF LILCILSATI
      101 NAGAVAIUTA AIVKMAIPSL MFDAGTVAAL IMASCLILV SGRYRALDRV
      151 SKIIIVTLSI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW
      201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL
50     251 GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
      301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIVVAGS GLAVIEWFDG
      351 VMANLLKFAM IAAVFSAPVF AWLNRYLVKG DEKHKLTSKM NALALAGLIY
      401 LTGFTVLELL NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N.meningitidis*:

-283-

	orf53.pep			10	20	30
				VSGRYRALDRVSKI	IIVTLSIATLAAAGIA	
	orf53a	AAIVKMAIPSLMFDAGTVAALIMASCLII	ILVSGRYRALDRVSKI	IIVTLSIATLAAAGIA		
5		110	120	130	140	150
	orf53.pep		40	50	60	70
			MSRGMQMOSDFIEPTPWTLAGLGLIALMGWMPAP	IEISAINSLWVTEKQRINPSEYRDG		
10	orf53a	170	180	190	200	210
			MSRGMQMOSDFIEPTPWTLAGLGLIALMGWMPAP	IEISAINSLWVTEKQRINPSEYRDG		
	orf53.pep		100	110	120	130
			IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQ	MAGGKYNGQLINMYA		
15	orf53a	230	240	250	260	270
			IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQ	MAGGKYIGQLINMYAVTIGGWSRPLV		
	orf53a	290	300	310	320	330
20			AFIAFACMYGTTITVVDGYARAIAEPVRLLRGKDK	TGNAEFFAWNIWVAGSGLAVIFWFD		

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

1	ATGTCCGAAC	AACATATTTT	GA	CTTGAAAA	AGTAAAATCA	ACGCATTGGG
51	ACCGGGGATT	ATGATGGCTT	CG	CGCGCGCGT	CGCGCGTTCG	CACCTGATTG
101	CCTCGACGCA	GGCGGGCGCG	CT	TACGGCT	GGCAGATCGC	GCTCATCATC
151	ATCCTGACCA	ACCTCTCAA	AT	ACCCGTTT	TTCCGCTTCA	GCGCGCATTA
201	CACGCTGGAC	ACGGGCAAGA	GC	CTGATTGA	AGGTATGCC	GAGAAAAGCC
251	GCGTTTATTT	GTGGGTATTC	CT	GATTTTGT	GCATCCTCTC	CGCCACGATT
301	AACGCGGGCG	CGGTCGCCAT	TG	TAAACCGCC	GCCATCGTCA	AAATGGCGAT
351	TCCCTCGCTG	ATGTTTGTATG	CC	GGCACGGT	TGCCGCCTTG	ATTATGGCAT
401	CCTGCCTGAT	TATTTTGGTG	AG	CGGACGTT	ACCGCGCTTT	GGATCGCGTT
451	TCCAAAATCA	TCATCGTTAC	TT	TGAGTATC	GCCACGCTTG	CCGCCGCCGG
501	CATCGCTATG	TCGCGCGGTA	TG	CAGATGCA	GTCCGATTTT	ATCGAGCCGA
551	CACCGTGGAC	GCTTGCCGGT	TT	GGGCTTCC	TGATCGCGCT	GATGGGCTGG
601	ATGCCCGCGC	CGATTGAAAT	TT	CCGCCATC	AATTCTTTGT	GGGTAAACCGA
651	AAAACAACGC	ATCAATCCTT	CC	GAATACCG	CGACGGGATT	TTTGATTTC
701	ACGTCGGTTA	TATCGCCAGT	GC	GGTTTGGG	CTTTGGTTTT	CCTTGCACTG
751	GGCGCGTTTG	TGCAATACGG	CA	ACGGCGAA	GCAATGCAGA	TGGCGGGCGG
801	CAAATATATC	GGGCAATTGA	TC	AATATGTA	CGCCGTTACC	ATCGGCGGCT
851	GGTCGCGCCC	GCTGGTGGCG	TT	TATCGCGT	TGCTGTAT	GTACGGCAGC
901	ACGATTACCG	TTGTGGACGG	CT	ATGCCCGT	GCCATGCGG	AACCGGTGCG
951	CCTGCTGCGC	GGAAAAGACA	AA	ACGGGCAA	CGCCGAATTC	TTTGCTGGA
1001	ATATTTGGGT	GGCGGGCAGC	GG	TTTGGCGG	TGATTTCTG	GTTTGACGGC
1051	GTAATGGCGA	ATCTGCTCAA	ATT	TGCGATG	ATTGCCGCTT	TTGTGTCCGC
1101	CCCTGTGTTT	GCCTGGCTGA	ATT	ACCGTTT	GGTCAAAGGT	GATGAAAAAC
1151	ACAAACTCAC	ATCAGGTATG	AAT	GCCCTTG	CATTGGCAGG	CTTGATTAT
1201	CTGACCGGTT	TTACCGTTTT	GTT	CTTATTG	AATTGGCGG	GAATGTTCAA
1251	ATGA					

This encodes a protein having amino acid sequence <SEQ ID 482>:

1	MSEQHISTWK	SKINALGPGI	MMASAAVGGG	HLIASTQAGA	LYGWQIALII
51	ILTNLFKYPF	FRFSAHYTLD	TGKSLIEGYA	EKSRVYLWVF	LILCILSATI
101	NAGAVAIVTA	AIWKMAIPSL	MFDAGTVAAL	IMASCLIIIV	SGRYRALDRV
151	SKIIIVTSLI	ATLAAAGIAM	SRGMQMOSDF	IEPTPWTLAG	LGFLIALMGW
201	MPAPIEISAI	NSLWVTEKQR	INPSEYRDGI	FDNFVGYIAS	AVLALVFLAL
251	GAFVQYGNGE	AVQMAGGKYI	GQLINMYAVT	IGGWSRPLVA	FIAFACMYGT
301	TITVVDGYAR	AIAEPVRLLR	GKDKTGNAEF	FAWNIWVAGS	GLAVIFWFDG
351	VMANLLKFAM	IAAFVSAPVF	AWLNRYLVKG	DEKHKLTSGM	NALALAGLIY
401	LTGFTVLFL	NLAGMFK*			

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

	orf53a.pep	10	20	30	40	50	60
60		MSEQHISTWKS	SKINALGPGIMMASAAVGGSH	LIAS	TQAGALYGWQIALII	ILTNLFKYPF	
	orf53-1	10	20	30	40	50	60
65		MSEQHISTWKS	SKINALGPGIMMASAAVGGSH	LIAS	TQAGALYGWQIALII	ILTNLFKYPF	
		70	80	90	100	110	120

-284-

5	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIVKMAIPSL
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIVKMAIPSL
10	orf53a.pep	130 140 150 160 170 180 MFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRGMQMQSDF
	orf53-1	MFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRGMQMQSDF
15	orf53a.pep	190 200 210 220 230 240 IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS
	orf53-1	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS
20	orf53a.pep	250 260 270 280 290 300 AVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
	orf53-1	AVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
25	orf53a.pep	310 320 330 340 350 360 TITVVDGYARAIAPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM
	orf53-1	TITVVDGYARAIAPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM
30	orf53a.pep	370 380 390 400 410 IAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLLNLAGMFKX
	orf53-1	IAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLLNLAGMFKX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N. gonorrhoeae*:

40	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIA	91
45	orf53.pep	MSRGMQMQSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
	orf53ng	MSRGMQMQPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
50	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQMAGGKYNGQLINMYA	139
	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQMGGGKYIGQLINMYAVTIGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

55	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMQP
	101	DFIEPTPWT LAGLGFLIALM GWMPAPIEIS AINSLWVTEK QRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
	201	VTIGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRGKDKTARP
	251	IVLLEKLGR HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
	51	CGCCACGATT AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA
	101	AAATGGCGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG

-285-

5
 10
 15

```

151 ATTATGGCAT CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT
201 GGATCGTGTT TCCAAAATCA TCATTGTTAC TTTGAGCATC GCCACGCTTG
251 CCGCCGCCGG CATCGCTATG TCGCGCGGTA TGCAGATGCA GCCCCGATTTT
301 ATCGAGCCGA CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT
351 GATGGGCTGG ATGCCCGCGC CGATCGAAAT TTCCGCCATC AATTCTTTGT
401 GGGTAACCGA AAAACAACGC ATCAATCCTT CTGAATACCG CGACGGGATT
451 TTCGATTTC ACGTCGGTTA TATCGCcagT GCGGTTTTTG CTTTGGTTTT
501 CCTTGCACTG GCGCGCTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA
551 TGGCGGGCGG CAAATATATC GGGCAATTGA TTAATATGTA TGCCGTAACC
601 ATCGGCGGCT GGTCTCGTCC GCTGGTGCGG TTTATCGCGT TTGCCTGTAT
651 GTACGGCACG ACGATTACCG TTGTGGACGG TTATGCGCGT GCCATTGCCG
701 AACCCGTGCG CCTGCTGCGG GGCAGGATA AAACCGGCAA CGCCGAGTTG
751 TTTgccTGGA ATATTTGGGT GGCGGGCAGC GGTTTGGCGG TGATTTTCTG
801 GTTTGACggc gcaaTGGcGg AActgcTCAA ATTTGCGATG ATtgccgcCT
851 TTGTGTCCGC CCCTGTGTTC GCCTGGCTCA ACTACCGCCT CGTCAAAGGG
901 GACAAACGCC ACAGGCTTAC CGCCGGTATG AACGCCCTTG CCATTGTCCG
951 CCTGCTCTAC CTGGCCGGGT TTGCCGTTTT GTTCCTGTTG AACCTTACCG
1001 GACTTTTGGC ATAG
  
```

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

20
 25

```

1 ..KKSCVYLWVF LILCIASATI NAGAVAIUTA AIVKMAIPSL MFDAGTVAAL
51 IMASCLIIIV SGRYRALDRV SKIIIVTSLI ATLAAAGIAM SRGMQMOPDF
101 IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEKOR INPSEYRDGI
151 FDFNVGYIAS AVLALVFLAL GAFVQYNGE AVQMAGGKYI GOLINMYAVT
201 IGGWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL
251 FAWNIIWVAGS GLAVIFWFDG AMAELLKFAM IAFVSAFVF AWLNRYLVKG
301 DKRHRLTAGM NALAIVGLLY LAGFAVLFL NLTGLLA*
  
```

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

30
 35
 40
 45
 50
 55
 60
 65

```

        60      70      80      90      100     110
orf53-1.pep  ILTNLFKYPFFRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIUTA
orf53ng-1      :||| ||||| ||||| ||||| ||||| |||||
               10      20      30
               KKSCVYLWVFLILCIASATINAGAVAIUTA

        120     130     140     150     160     170
orf53-1.pep  AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIAM
orf53ng-1    AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIAM
               40      50      60      70      80      90

        180     190     200     210     220     230
orf53-1.pep  SRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKORINPSEYRDGI
orf53ng-1    SRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKORINPSEYRDGI
               100     110     120     130     140     150

        240     250     260     270     280     290
orf53-1.pep  FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA
orf53ng-1    FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA
               160     170     180     190     200     210

        300     310     320     330     340     350
orf53-1.pep  FIAFACMYGTTITVVDGYARAIAEPVRLLRGDKDTGNAEFFAWNIIWVAGSGLAVIFWFDG
orf53ng-1    FIAFACMYGTTITVVDGYARAIAEPVRLLRGRDKDTGNAELFAWNIIWVAGSGLAVIFWFDG
               220     230     240     250     260     270

        360     370     380     390     400     410
orf53-1.pep  VMANLLKFAMIAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALAGLIYLTGFTVLFL
orf53ng-1    AMAELLKFAMIAAFVSAPVFAWLNRYLVKGDKRHRLTAGMNALAIVGLLYLAGFAVLFL
               280     290     300     310     320     330

orf53-1.pep  NLAGMFKX
orf53ng-1    NLTGLLAX
  
```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 5 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

      1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
    51  TGCGCTTGCC GGCTTGTGTTT TTGTCCGCGC ACAATCCGAA CGCGAGTGGG
  101  TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG
    151  CTGCCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT
    201  GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTCTG
    251  TCCGTTTCTG CCGAACTAT CTGGCGCACG AATCCGAACC GGACAGGCC
    301  GTCCGCCT..
  
```

15 This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

      1  ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVS AW QEKKGEKQAE
    51  LPEIKDGMFD FPELALMLFH AVKTAVYWLF VGVVRFRCRNY LAHESEPD RP
   101  VPP..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

  20      1  ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
    51  GTTTTTTGTC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
   101  CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
   151  GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
   201  CAAAACGGCA GTGTATTGGC TGTGTCGG TGTCGTCCGT TTCTGCCGAA
   251  ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
   301  GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
   351  AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
   401  AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
   451  ATCCCATTCG ACCGAGATAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
   501  AATTTGCCCC GTCCGTCCGG TTTTAAAGA AATCACTTTG GAAGAAGCAA
   551  CGCGTGCTTT AAACAGCGCG GCTTAAAGGG AAACGAAAAA ACGCTATATC
   601  GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCGCA
   651  TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
   701  AACGCACGTA TTCCCATATG TTCGATGCGG ACAAGAAGC GTTTTCCGAG
   751  TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
   801  CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACCGTC
   851  ATGCAGGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
   901  CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCCGCCCGC
   951  CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
  1001  CGCGAATTTT TCGCCTGATT CCGGAAAGTC AGACGGTTGT CGGGAAACGG
  1051  GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTACGG AAACCGTTTC
  1101  GTCTGTGGA TACGGCGGTC CGGTTTATGA TGAACTGCC GATATCCATA
  1151  TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
  1201  CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC
  1251  GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAGGATTG GAGCAGGTGC
  1301  AACGCAGCCG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTTGAAT
  1351  GGAGGTGGC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG
  1401  TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAAACC GAAGCGTTCG
  1451  GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
  1501  CCGTCCTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCCATC
  1551  TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
  1601  TGCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAAGTGTG
  1651  GAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
  1701  CAAGGTGTG GATTCTTATT CCGGCCCGGT AATTACGCGT TATGAAATCG
  1751  AATCCGATGT CGGCGTGCGC GGCAATTCCG TTCTGAATCT GGAAAAAGAT
  1801  TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCT AAACCATCCC
  1851  CGGCAAAACC TGCATGGGTT TGGAATTCC GAACCCGAAA CGCCAAATGA
  
```

-287-

1901 TACGCCTGAG CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAAATCC
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
 2001 CGACTTGGGA AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGGTTCGG
 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
 2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAA TGCTGGAATT
 2151 GAGCATTAC GAAGGCATCC CGCACCTGCT CGCCCCTGTC GTTACCGATA
 2201 TGAAGCTGGC GGCAAACGCG CTGAAGTGGT GTGTTAACGA AATGGAAAAA
 2251 CGCTACCGCC TGATGAGCTT TATGGCGCTG CGTAATCTTG CGGGCTTCAA
 2301 TCAAAAAATC GCCGAAGCGG CAGCAAGGGG AGAAAAAATC GGCAATCCGT
 2351 TCAGCCTCAC GCGCGACGAT CCCGAACCTT TGGAAAAACT GCCGTTTATC
 2401 GTGGTCGTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
 2451 AATCGAAGAA CTGATTGCCG GCCTCGCCCA AAAAGCCCGC GCGCGAGGCA
 2501 TCCATTGAT TCTTGCCACA CAACGCCCCA GCGTCGATGT CATCACGGGT
 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCAGG TGTCCAGCAA
 2601 AATCGACAGC CGCACGATTC TCGACCAAT GGGCGCGGAA AACCTGCTCG
 2651 GTCAGGGCGA TATGCTGTTT CTGCTGCCGG GTACTGCCTA TCCGCAAGCG
 2701 GTTCACGGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTCAATA
 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG
 2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGACGAAACC
 2851 GATCGGATGT ACGACGAGGC CGTATCCGT GTCTTGAAAA CGCGCAAGGC
 2901 CAGCATTTCC GGCCTACAGC GCGCCTTGCG TATCGGCTAC AACCGCGCCG
 2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25 1 MFWIVLIVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
 51 DGMPDFPELA LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPPAS
 101 ANRADVPTAS DGYSDSGNGT EEAEETEEAEA AEEEEADTED IATAVIDNRR
 151 IPFORSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE
 30 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQKG QAEAKSPDVS
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR
 351 DVEMPSETEN VETETVSSVG YGGPVYDETA DIHIEEPAAP DAWVVEPPEV
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPPSGF EQVQSRSAE TDHLADDVLN
 451 GGWQEETAAI ADDGSEGAEE RSSGQYLSET EAFGHDSQAV CPFENVPSER
 35 501 PSCRVSDETA DEGAFFSEET GAVSEHLPTT DLLLPPLFNP EATQTEELL
 551 ENSITIEEKL AEFKVKVKVV DSYSGPVITR YEIEPDVGVV GNSVLNLEKD
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 651 KLTLAGQDI TGQPVVTDLG KAPHLVAGT TSGSKSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMLAANA LNWCVNEMEK
 40 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTDD PEPLEKLPFI
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPVSDVITG
 851 LIKANIPTRI AFOVSSKIDS RTILDQMGAE NLLGQDMLF LLPGTAYPQR
 901 VHGAFAFASDEE VHRVVEYLKQ FGEFDYVDDI LSGGGSEELP GIGRSGDDET
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
 45 1001 HNGNRTILVP LDNA*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

50 *meningitidis*:

orf58.pep 10 20 30 40 50 60
 LRETAYVLDSFDRYFVVALAGLFFVRAQSEREWMREVS AWQEKKG EKQALPEIKDGMPD
 orf58a MFWIVLIVILLALAGLFFVRAQSEREWMREVS AWQEKKG EKQALPEIKDGMPD
 55 10 20 30 40 50
 orf58.pep 70 80 90 100
 FPALALMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDRPVPP
 orf58a FPALALMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDRPVPPASANRADVPTASDGYSD
 60 60 70 80 90 100 110

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

1  ATGTTTGGGA TAGTTTGGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTGTGC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101  CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
5  151  GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201  CAAAACGGCA GTGTATTGCG TGTGTGTCGG TGTCGTCCGT TTCTGCCGAA
251  ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301  GCAAATCGTG CCGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
10  351  AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
401  AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451  ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
501  AATTTGCCCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551  CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGGTATATC
601  GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
15  651  TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701  AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG
751  TCTGCGGATT ACGGATTGTA GCCGTATTTT GAGAAGCAGC ATCCGTATGC
801  CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCCGTC
851  ATGCAGGGCA GGGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
20  901  CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCNGCCGCCG
951  CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGTAGG
1001  CGCGGATTTT GCGCCTGATT CCGGAAAGTC GGACGGTTGT CGGGAAACGG
1051  GATGTCGAAA TGCCGTCTGA AACCGAAAAA GTTTTACCGG AAANTGTTTC
1101  GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAAACTGCC GATATCCATA
25  1151  TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
1201  CCGAAAGTTC CCATGCCCGC AATNGATATT CCGCCGCCGC CTCCCGTATC
1251  GGAAATCTAC AACCCTACCT ATGAACCGCC GGCAGGATTC GAGCAGGTGC
1301  AACGCAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTTGAAT
1351  GGAGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTGGGG
30  1401  TGTGGCAGAG CGGTCAAGCG GGCAATATTT GTCGGAACC GAAGCGTTTCG
1451  GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
1501  CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
1551  TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
35  1601  TGCCGCGGCT GTTCAATCCC GGGCGACGC AAACCGAAGA AGANCTGTTG
1651  GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701  CAAGTTGTC GATTTTATT CCGCCCCGT GATTACGCGT TATGAAATCG
1751  AACCCGATGT CCGCGTGCGC GGCAATTCCG TTCTAAATCT GGAAAAAGAN
1801  TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCT AAACCATCCT
40  1851  CGGCAAAACC TGTATGGGTT TGGAACCTCC GAACCCGAAA CGCCAAATGA
1901  TACGCTGAG CGAAATCTC AATTGCCCCG AGTTTGCCGA ATCCAAATCC
1951  AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCTGAAC
2001  CGACTTGGGC AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGTTTCGG
2051  GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
45  2101  GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151  GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCCGTGC GTTACCGATA
2201  TGAAGCTGGC GGCAACGCG CTGAACCTGT GTGTTAACGA AATGGAAGAAA
2251  CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGTNTCAA
2301  TCAAAAAATC CGCGAAGCCG CAGCAAGGGG GGAGAAAAATC GGCAACCCGT
50  2351  TCAGCCTCAC GCCCGACAAT CCCGAACCTT TGGANAAATT GCCGTTTATC
2401  GTGGTCGTGG TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451  AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2501  TCCATCTTAT CCTTGCCACA CAACGCCCCA GTGTCGATGT CATCACGGGT
2551  CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
55  2601  AATCGACAGC CGCACGATTC TTGACCAAT GGGTGCGGAA AACCTGCTCG
2651  GGCAGGGCGA TATGCTGTTC CTGCCGCCGG GTACGGCCTA TCCGCAGCGC
2701  GTTCACGGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA
2751  TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
2801  GTATGTCCGA CGATTTGCTG GGAATCAGCC GGAGCGCGCA CGGCGAAACC
60  2851  GATCCGATGT ACGACGAGGC CGTGTCNGTT GTTTTGAAAA CGGCGAAAGC
2901  GAGCATTTCT GGCGTGAGC GCGCATTCG TATCGGCTAT AATCGCGCCG
2951  CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
3001  CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

65  1  MFWIVLIVIL LLAGLGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK
51  51  DGMPDFPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS
101  101  ANRADVPTAS DGYSDSNGT EEAETEEAEA AEEEEADTED IATAVIDNRR
151  151  IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
201  201  DAFEKNETAV PKVRVSDTPM EGLQIIIGLDD PVLQRTYSRM FDADKEAFSE

```

-289-

5 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQGKG QAEAKSPDVS
 301 QGQSVSDGTA VRDAXRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
 351 DVEMPSETEN VFTEXVSSVG YGXPVYDETA DIHIEEPAAP WDAWVVEPPEV
 401 PKVPM PAXDI PPPPVSEIY NRTYEPPAGF EQVQRSRIAE TDHLADDVLN
 451 GGWQEETA AI ANDGSEGVAE RSSGQYLSET EAFGHDSQAV CPFENVPSER
 501 PSRRAXDTEA DEGAFOSEET GAVSEHLPTT DLLLPPLFNP GATQTEEXLL
 551 XNSITIEEKX AEFKVKVKVV DSYSGPVITR YEIEPDVGV R GNSVLNLEKX
 601 LARSLGVASI RVVETILGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 10 651 KLTLALGQDI TGQPVVTDLG KAPHLVAGT TGSKGKSVGN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGIPHLAPV VTDMKLAANA LNWCVNEMEK
 751 RYRLMSFMGV RNLAGXNQKI AEAARGEKI GNPFSLTPDN PEPLXKL PFI
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPVSDVITG
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGA E NLLGQGDMLF LPPGTAYPQR
 901 VHGAFA SDEE VHRVVEYLKQ FGEPDYDDX LSGGMSDDL GISRSGDGET
 15 951 DPMYDEAVSV VLKTRKASIS GVQALRIGY NRAARLIDQ EAEGIVSAPE
 1001 HNGNRTILVP XDNA*

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

20	orf58a.pep	10	20	30	40	50	60
	orf58a.pep	MFWIVLIVILLALLAGLFFVRAQSEREWREVS AWQEKKGKQAE LPEIKDGM PDPFPELA					
	orf58-1	MFWIVLIVILLALLAGLFFVRAQSEREWREVS AWQEKKGKQAE LPEIKDGM PDPFPELA					
		10	20	30	40	50	60
25	orf58a.pep	70	80	90	100	110	120
	orf58a.pep	LMLFHAVKTAVYWL FVGVRFCRNYLAH ESEPDRPVPPAS ANRADVPTASD GYSDSGNGT					
	orf58-1	LMLFHAVKTAVYWL FVGVRFCRNYLAH ESEPDRPVPPAS ANRADVPTASD GYSDSGNGT					
		70	80	90	100	110	120
30	orf58a.pep	130	140	150	160	170	180
	orf58a.pep	EEAETEEAEAE EEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPV RPVFKEITL					
	orf58-1	EEAETEEAEAE EEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPV RPVFKEITL					
		130	140	150	160	170	180
35	orf58a.pep	190	200	210	220	230	240
	orf58a.pep	EEATRALNSAALRETKKRYIDAF EKNETAVPKVRVSDT PMEGLQIIGLDDPVLQRTYSRM					
	orf58-1	EEATRALNSAALRETKKRYIDAF EKNETAVPKVRVSDT PMEGLQIIGLDDPVLQRTYSHM					
		190	200	210	220	230	240
40	orf58a.pep	250	260	270	280	290	300
	orf58a.pep	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS					
	orf58-1	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS					
		250	260	270	280	290	300
45	orf58a.pep	310	320	330	340	350	360
	orf58a.pep	QGQSVSDGTAVRDAXRRVSVNLKEPNKATVSAE ARISRLI PESRTVVGKRDVEMPSETEN					
	orf58-1	QGQSVSDGTAVRDARRVSVNLKEPNKATVSAE ARISRLI PESQTVVGKRDVEMPSETEN					
		310	320	330	340	350	360
50	orf58a.pep	370	380	390	400	410	420
	orf58a.pep	VFTEXVSSVG YGXPVYDETA DIHIEEPAAPDAWVVEPPEV PKVPM PAXDI PPPPVSEIY					
	orf58-1	VFTETVSSVG YGGPVYDETA DIHIEEPAAPDAWVVEPPEV PKVPM TAI DIQ PPPPVSEIY					
		370	380	390	400	410	420
55	orf58a.pep	430	440	450	460	470	480
	orf58a.pep	NRTYEPPAGFEQVQRSRIAE TDHLADDVLNGGWQEETA AIANDGSEGVAE RSSGQYLSET					
	orf58-1	NRTYEPPSGFEQVQRSRIAE TDHLADDVLNGGWQEETA AIADDGSEGA AE RSSGQYLSET					
		430	440	450	460	470	480
60	orf58a.pep	490	500	510	520	530	540
	orf58a.pep	EAFGHDSQAVCPFENVPSERPSRRAXDTEA DEGAFOSEETGAVSEHLPTT DLLLPPLFNP					
	orf58-1	EAFGHDSQAVCPFENVPSERPSRVSDTEA DEGAFFSEETGAVSEHLPTT DLLLPPLFNP					
		490	500	510	520	530	540
65							
70							

-290-

		550	560	570	580	590	600
	orf58a.pep	GATQTEEXLLXNSITIEEKXAEFKVKVVDVSYSGPVITRYEIEPDVGVRGNSVLNLEKX					
5	orf58-1	EATQTEEELENSITIEEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVRGNSVLNLEKD					
		550	560	570	580	590	600
		610	620	630	640	650	660
10	orf58a.pep	LARSLGVASIRVVETILGKTCMGLLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
	orf58-1	LARSLGVASIRVVETIPGKTCMGLLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
		610	620	630	640	650	660
		670	680	690	700	710	720
15	orf58a.pep	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
	orf58-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
		670	680	690	700	710	720
		730	740	750	760	770	780
20	orf58a.pep	EGIPHLLAPVVTDMDKLAANALNWCNVNEMEKRYRLMSFMGVRNLAGXNQKIAEAAAARGEKI					
	orf58-1	EGIPHLLAPVVTDMDKLAANALNWCNVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAAARGEKI					
		730	740	750	760	770	780
25		790	800	810	820	830	840
	orf58a.pep	GNPFSLTPDNPEPLXKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
30		790	800	810	820	830	840
		850	860	870	880	890	900
	orf58a.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR					
35	orf58-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
		850	860	870	880	890	900
		910	920	930	940	950	960
40	orf58a.pep	VHGAFASDEEVHRVVEYLKQFGEPTYVDDXLGGMSDDLGLISRSRGDGETDPMYDEAVSV					
	orf58-1	VHGAFASDEEVHRVVEYLKQFGEPTYVDDILSGGSEELPGIGRSGDDETDPYDEAVSV					
		910	920	930	940	950	960
		970	980	990	1000	1010	
45	orf58a.pep	VLKTRKASISGVQALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPXDNAX					
	orf58-1	VLKTRKASISGVQALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

	orf58.pep	ALMLFHAVKTAVYWLFVGVVRFRCRNYLAHESEPDRPVPP	103
55	orf58ng	SEPDRPVPPASANRADVPTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

	1	SEPDRPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETAAE	AAAAAADTE
60	51	DIATAVIDNR	RIPFDRSIAE	GLMQSESKTS	PVRPVFKEIT	LEEATRALSS
	101	AALRETKKRY	IDAFEKNGTA	VPKVRVSDTP	MEGLQIIGLD	DPVLQRTYSR
	151	MFDADKEAFS	ESADYGFEPY	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK
	201	GQAEAKSPDV	SQGQSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL
	251	IPESRTVVGK	RDVEMPSETE	NVFTETVSSV	GYGGPVYDEA	ADIHIEEPAA
	301	PDWVVEPPE	VPEVAVPEID	ILPPPPVSEI	YNRTYEPPAG	FEQAQRSRIA

351	ETDHLAADVL	NGGWQEETAA	IADDGSEGAA	ERSSGOYLSE	TEAFGHDSQA
401	VCPFEDVPSE	RPSCRVSDE	ADEGAFQSEE	TGAVSEHLPT	TDLLLPPLFN
451	PEATQTEEL	LENSITIEEK	LAEFKVKVKV	VDSYSGPVIT	RYEIEPDVGV
501	RGNSVLNLEK	DLARSLGVAS	IRVVETIPGK	TCMGLELPNP	KRQMIRLSEI
551	FNSPEFAESK	SKLTALGQD	ITGQPVVTDL	GKAPHLVAG	<u>TTGSGKSVGV</u>
601	NAMILSMLFK	AAPEDVRMIM	IDPKMLELSI	YEGITHLLAP	VVTDMKLAAN
651	ALNWCVNEME	KRYRLMSFMG	VRNLAGFNQK	IAEAAARGEK	IGNPFSLTPD
701	DPEPLEKLPF	IVVVVDEFAD	LMMTAGKKIE	ELIARLAQKA	RAAGIHLILA
751	TQRPSVDVIT	GLIKANIPTR	IAFQVSSKID	SRTILDQMG	ENLLGQDML
801	FLPPGTAYPO	RVHGAFASDE	EVHRVVEYLK	QFGEPTYDD	ILSGGGSEEL
851	PGIGRSGDGE	TDPMYDEAVS	VVLKTRKASI	SGVQRALRIG	YNRAARLIDQ
901	MEAEGIVSAP	EHNGNRTILV	PLDNA*		

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng:	467	IEEKLAEFKVKVKVDSYSGPVITRYEIEPDVGRGNSVLNLEKDLARSLGVASIRVVET	526
		+E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE	
FtsK:	868	VEARLADFRIKADVNNYSPGPVITRFEINLAPGVKAARISNLSRDLARSLSTVAVRVVEV	927
ORF58ng:	527	IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL	586
		IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL	
FtsK:	928	IPGKPYVGLELPNKKRQTVYLREVLDAKFRDNPSPLTVVLGKDIAEPVADLAKMPHL	987
ORF58ng:	587	LVAGTTGSGKSVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMMK	646
		LVAGTTGSGKSVGNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK	
FtsK:	988	LVAGTTGSGKSVGNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVTDMMK	1047
ORF58ng:	647	LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP--	704
		AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +	
FtsK:	1048	DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPPIPDYWKPGDSMDAQH	1107
ORF58ng:	705	--LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL	762
		L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL	
FtsK:	1108	PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL	1167
ORF58ng:	763	IKANIPTRIAFQVSSKIDSRITILDQMGAEENLLGQDMLFLPPGTAYPQRVHGAFASDEEV	822
		IKANIPTRIAF VSSKIDSRITILDQ GAE+LLG GDML+ P + P RVHGAF D+EV	
FtsK:	1168	IKANIPTRIAFTVSSKIDSRITILDQAGAESLLGMGMPLYSGPNSTLPVRVHGAFVRDQEV	1227
ORF58ng:	823	HRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG	882
		H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG	
FtsK:	1228	HAVVQDWKARGRPQYVDGITSDSESEGGAG-GFDGAELDPLFDQAVQVTEKRRKASISG	1286
ORF58ng:	883	VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP	921
		VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P	
FtsK:	1287	VQRQFRIGYNRAARLIEQMEAQGIQVSEQHNGNREVLAP	1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGTTTTGGA	TAGTTTTGAT	CGTTATgtg	TTGCTTGC	CGCCTTGC
51	GTTTTTTGTC	CGCGCACAA	CCGAACGCG	GTGGATGCG	GAGGTTTCTG
101	CGTGGCAGGA	AAAGAAAGG	GAAAAACAG	CGGAGCTGC	TGAAATCAAA
151	GACGGTATGC	CCGATTTTCC	CGAGTTTTC	CTGATGCTTT	TCCATGCCGT
201	CAAAACGGCA	GTGTATTGCG	TGTTTGTGCG	TGTCGTCGGT	TTCTGCCGAA
251	ACTATCTGGC	GCACGAATCC	GAACCGGACA	GGCCCGTTCC	GCCTGCTTCT
301	GCAAACCGTG	CGGATGTTCC	GACCGCATCC	GACGGGTATT	CAGACAGTGG
351	AAACGGGACG	GAAGAAGCGG	AAACGGGAAG	AGCAGAAGCT	GCGGAGGAAG
401	AGGCTGCCgA	TACgGAAGAC	ATTGCAACTG	CCGTATCGA	CAACCGCCGC
451	ATCCcatTCG	ACCGGAGTAT	TGCTGAAGGG	TTGATGCACT	CTGAAAGCAA
501	AACCTTCGCC	GTCCGTCCGG	TTTTTAAGGA	AATCACTTTG	GAAGAAGCAA
551	CGCGTGCTTT	AAGCAGCGCG	GCTTTAAGGG	AAACGAAAAA	ACGCTATATC
601	GATGCATTTG	AGAAAAACGG	AACAGCCGTC	CCCAAAGTAC	GCCTGTCCGA
651	TACCCCGATG	GAAGGGCTGC	AGATTATCGG	TTTGGACGAC	CCTGTGCTTC
701	AACGCACGTA	TTCCCGTATG	TTTGATGCGG	ACAAAGAAGC	GTTTTCGAG
751	TCTGCGGATT	ACGGATTTGA	GCCGTATTTT	GAGAAGCAGC	ATCCGTCTGC

	801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCACG	GAATGCGCCG	TTCCGCCGTC
	851	ATGCAGGGCA	GGAGAAAGGG	CAGGCGGAGG	CAAAATCCCC	GGATGTTTCC
	901	CAAGGGCAGT	CCGTTCAGAG	CGGCACAGCC	GTCCGCGATG	CCCGCCGCCG
5	951	CGTTTCCGTC	AATTTGAAAG	AACCGAACAA	GGCAACGGTT	TCTGCGGAGG
	1001	CGCGGATTTT	GCGCCTGATT	CCGGAAGTTC	GGACGGTTGT	CGGGAAACGG
	1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAT	GTTTTACGG	AAACCGTTTC
	1101	GTCTGTGGGA	TACGCGGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
	1151	TTGAAGAGCC	TGCCGCGCCC	GATGCTTGGG	TGGTCGAACC	ACCCGAAGTG
10	1201	CCGGAGGTAG	CCGTACCCGA	AATCGATATT	CTGCCGCCGC	CTCCCGTATC
	1251	GGAAATCTAC	AACCGTACCT	ATGAGCCGCC	GGCAGGATTC	GAGCAGGCGC
	1301	AACGCGCCG	CATTGCCGAA	ACCGACCATC	TTGCCGCTGA	TGTTTTGAAT
	1351	GGAGGTTGGC	AGGAGGAAAC	CGCCGCTATT	GCAGATGACG	GCAGTGAGGG
	1401	TGCGGCAGAG	CGGTCAAGCG	GGCAATATCT	GTCGGAACCC	GAACGTTTCG
15	1451	GGCATGACAG	TCAGGCGGTT	TGTCCGTTTG	AAGATGTGCC	GTCTGAACGC
	1501	CCGTCTGCC	GGGTATCGGA	TACGGAAGCG	GATGAAGGGG	CGTTCCAATC
	1551	GGAAGAGACC	GGTGGCGTAT	CCGAACACCT	GCCGACAACC	GACCTGCTTC
	1601	TGCCCTCCGT	GTTCAATCCC	GAGGCGACGC	AAACCGAAGA	AGAAGCTTTG
	1651	GAAAAACAGCA	TCACCATCGA	AGAAAAATTG	GCGGAGTTCA	AAGTCAAGGT
20	1701	CAAGGTTGTC	GATTCTTATT	CCGCCCCCGT	GATTACGCGT	TATGAAATCG
	1751	AACCCGATGT	CGCGTGCGC	GGCAATTCCG	TTCTGAATTT	GGAAAAAGAC
	1801	TTGGCGCGTT	CGCTCGGCGT	GGCTTCCATC	CGCGTTGTTCG	AAACCATCCC
	1851	CGGCAAAACC	TGCATGGGTT	TGGAACCTCC	GAACCCGAAA	CGCCAAATGA
	1901	TACGCTGAG	CGAAATTTTC	AATTCGCCCC	AGTTTGCCGA	ATCCAAATCC
25	1951	AAGCTGACGC	TCGCGCTCGG	TCAGGACATT	ACCGGACAGC	CCGTCTTAAC
	2001	CGACTTGGGC	AAAGCACCGC	ATTGCTGGT	TGCCGGCACG	ACCGGTTTCG
	2051	GCAAATCGGT	GGGTGTCAAC	GCGATGATTC	TGTCTATGCT	TTTCAAAGCC
	2101	GCGCCGGAAG	ACGTGCGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
	2151	GAGCATTTAC	GAAGGCATCA	CGCACCTGCT	CGCCCCGTGC	GTTCCAGATA
30	2201	TGAAAGCTGGC	GGCAAACGCG	CTGAACTGGT	GTGTTAACGA	AATGGAAAAA
	2251	CGCTACCGCC	TGATGAGCTT	TATGGGCGTG	CGCAATCTTG	CGGGCTTCAA
	2301	CCAAAAAATC	GCCGAAGCCG	CAGCAAGGGG	AGAAAAAATC	GGCAATCCGT
	2351	TCAGCCTCAC	GCCCCGACAT	CCCGAACCTT	TGGAAAAACT	CCCGTTTATC
	2401	GTGGTCGTGG	TCGATGAGTT	TGCCGATTTG	ATGATGACGG	CAGGCAAGAA
35	2451	AATCGAAGAA	CTGATTGCGC	GCCTCGCCCA	AAAAGCCCGC	GCGGCAGGCA
	2501	TCCACCTTAT	CCTTGCCACA	CAACGCCCCA	GCGTCGATGT	CATCACGGGT
	2551	CTGATTAAGG	CGAACATCCC	GACGCGTATC	GCGTTCCAAG	TGTCCAGCAA
	2601	AATCGACAGC	CGCACGATTC	TCGACCAAAT	GGGCGCGGAA	AACCTGCTCG
	2651	GTCAGGGCGA	TATGCTGTTT	CTGCCGCCGG	GTAATGCCTA	TCCGACGCGC
40	2701	GTTACGGCG	CGTTTGCCCT	GGATGAAGAG	GTGCACCGCG	TGGTCGAATA
	2751	TCTGAAGCAG	TTTGCGGAGC	CGGACTATGT	TGACGATATT	TTGAGCGGCG
	2801	GCGGCAGCGA	AGAGCTGCCC	GGCATCGGGC	GCAGCGGCGA	CGGCGAAACC
	2851	GATCCGATGT	ACGACGAGGC	CGTATCCGTT	GTCCTGAAAA	CGCGCAAAGC
	2901	CAGCATTTTC	GGCGTACAGC	GCGCCTTGCG	CATCGGCTAC	AACCCGCGCC
45	2951	CGCGTCTGAT	TGACCAAATG	GAAGCGGAAG	GCATTGTGTC	CGCACCGGAA
	3001	CACAACGGCA	ACCGTACGAT	TCTCGTCCCC	TTGGACAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

	1	MFWIVLIVIV	LLALAGLFFV	RAQSEREWMR	EVSAWQEKKG	EKQALPEIK
	51	DGMPDFPEFS	LMLFHAVKTA	VYWLFVGVVR	FCRNYLAHES	EPDRPVPPAS
50	101	ANRADVPTAS	DGYSDSGNGT	EEAEETAAEA	AEEEEADTED	IATAVIDNRR
	151	IPFDRSIAEG	LMQSESKTSP	VRPVFKEITL	EEATRALSSA	ALRETKKRYI
	201	DAFEKNGTAV	PKVRVSDTPM	EGLQIIGLDD	PVLQRTYSRM	FDADKEAFSE
	251	SADYGFEPLY	EKQHPSAFSA	VKAENARNAP	FRRHAGQKEG	QAEAKSPDVS
	301	QGQSVSDGTA	VRDARRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
55	351	DVEMPSETEN	VFTETVSSVG	YGGPVYDEAA	DIHIEEPAAP	DAWVVEPEV
	401	PEVAVPEIDI	LPPPPVSEIY	NRTYEPPAGF	EQAQRSRIAE	TDHLAADVLN
	451	GGWQEETAAL	ADDGSEGAEE	RSSGQYLSET	EAFGHDSQAV	CPFEDVPSE
	501	PSCRVSDETA	DEGAFQSEET	GAVSEHLPTT	DLLLPLPLFP	EATQTEEELL
	551	ENSITIEEKL	AEFKVKVKVV	DSYSGPVITR	YEIEPDVGVR	GNFVNLLEKD
60	601	LARSLGVASI	RVVETIPGKT	CMGLELPNPK	QMIRLSEIF	NSPEFAESKS
	651	KLTLALQODI	TGQPVVTDLG	KAPHLVAGT	TGSGKSVGVN	AMILSMLFKA
	701	APEDVRMIMI	DPKMLELSIY	EGITHLLAPV	VTDMKLAANA	LNWCVNEMEK
	751	RYRLMSFMGV	RNLAGFNQKI	AEAAARGEKI	GNPFSLTDD	PEPLEKLPFI
	801	VVVVDEFADL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
65	851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPPGTAYPQR
	901	VHGAFADEE	VHRVVEYLKQ	FGEPTYVDDI	LSGGGSEELP	GIGRSGDGET
	951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAPE
	1001	HNGNRTILVP	LDNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

-293-

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALLAGLFFVRAQSEREWREVS	SAWQEKKGKQAE	LPEIKDGM	PDFPELA		
	orf58ng-1	MFWIVLIVIVLLALLAGLFFVRAQSEREWREVS	SAWQEKKGKQAE	LPEIKDGM	PDFPEFS		
5		10	20	30	40	50	60
	orf58-1.pep	LMLFHAVKTAVYWL	FVGVVRF	CRNYLAH	SEPD	RPVPPAS	ANRADVPTASDGYSDSGNGT
10	orf58ng-1	LMLFHAVKTAVYWL	FVGVVRF	CRNYLAH	SEPD	RPVPPAS	ANRADVPTASDGYSDSGNGT
		70	80	90	100	110	120
	orf58-1.pep	EEAETEEAAAEAAE	EAADTED	IATAVID	NRRIP	FD	RSIAEGLMPSESEISPV
15	orf58ng-1	EEAETEEAAAEAAE	EAADTED	IATAVID	NRRIP	FD	RSIAEGLMQSESKTSPV
		130	140	150	160	170	180
	orf58-1.pep	EEATRALNSAAL	RETKKRY	IDAF	EKNET	AVPKVR	SDTPMEGLQIIGLDDPVLQRTYSHM
20	orf58ng-1	EEATRALSSAAL	RETKKRY	IDAF	EKNGT	AVPKVR	SDTPMEGLQIIGLDDPVLQRTYSRM
		190	200	210	220	230	240
	orf58-1.pep	FDADKEAFSES	ADYGF	EPEYFEK	QHPSA	FS	SAVKAENARNAPFHRHAGQGKGQAEAKSPDVS
25	orf58ng-1	FDADKEAFSES	ADYGF	EPEYFEK	QHPSA	FS	SAVKAENARNAPFRRHAGQEGKGQAEAKSPDVS
		250	260	270	280	290	300
30	orf58-1.pep	QGQSVSDGT	AVRDARR	RVSVNL	KEPNK	ATVSAE	ARISRLIPESQTVVGKRDVEMPSETEN
	orf58ng-1	QGQSVSDGT	AVRDARR	RVSVNL	KEPNK	ATVSAE	ARISRLIPESRTVVGKRDVEMPSETEN
35		310	320	330	340	350	360
	orf58-1.pep	VFTETVSSV	GYGGP	VYDE	TADIH	IEEPA	APDAWVVEPPEV
40	orf58ng-1	VFTETVSSV	GYGGP	VYDE	AADIH	IEEPA	APDAWVVEPPEV
		370	380	390	400	410	420
	orf58-1.pep	NRTYEPP	SGFEQ	VQRS	RIAET	DHLAD	DLVNGGWEETA
45	orf58ng-1	NRTYEPP	SGFEQ	VQRS	RIAET	DHLAD	DLVNGGWEETA
		430	440	450	460	470	480
	orf58-1.pep	EAFGHDS	QAVCP	FENVP	SERPS	CRVSD	TEADGAFQSEETGAVSEHLPTD
50	orf58ng-1	EAFGHDS	QAVCP	FEDVP	SERPS	CRVSD	TEADGAFQSEETGAVSEHLPTD
		490	500	510	520	530	540
	orf58-1.pep	EATQTEEE	LLENS	ITIEE	KLAE	FKVKV	VDSYSGPVITRYEIEPDVGVRGNSVLNLEKD
55	orf58ng-1	EATQTEEE	LLENS	ITIEE	KLAE	FKVKV	VDSYSGPVITRYEIEPDVGVRGNSVLNLEKD
		550	560	570	580	590	600
	orf58-1.pep	LARSLG	VASIR	VVETI	PGKTC	MGLEL	PNPKRQMI
60	orf58ng-1	LARSLG	VASIR	VVETI	PGKTC	MGLEL	PNPKRQMI
		610	620	630	640	650	660
	orf58-1.pep	TGQPVVT	DLGKAP	HLLV	AGTTG	SGKSV	GVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
65	orf58ng-1	TGQPVVT	DLGKAP	HLLV	AGTTG	SGKSV	GVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
		670	680	690	700	710	720
	orf58-1.pep	TGQPVVT	DLGKAP	HLLV	AGTTG	SGKSV	GVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
70	orf58ng-1	TGQPVVT	DLGKAP	HLLV	AGTTG	SGKSV	GVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
		670	680	690	700	710	720

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAAARGEKI					
	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAAARGEKI					
5		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
10		790	800	810	820	830	840
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
15		850	860	870	880	890	900
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
20		910	920	930	940	950	960
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
25		970	980	990	1000	1010	
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

30 Furthermore, ORF58ng-1 shows significant homology to the *E. coli* protein FtsK:

	sp P46889 FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli] >gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] >gi 1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329
35	Score = 576 bits (1469), Expect = e-163 Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)
	Query: 556 IEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 615 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE
40	Sbjct: 868 VEARLADFRKADVVNYSPPGPVITRFELNLAGVKAARISNLSRDLARSLSTVAVRVVEV 927
	Query: 616 IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGPQVVTDLGKAPHL 675 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL
45	Sbjct: 928 IPGKPYVGLLEPNKKRQTVYLREVLDAKFRDNPSPLTVVLGKDIAAGEPVVADLAKMPHL 987
	Query: 676 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK
	Sbjct: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVTDMK 1047
50	Query: 736 LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAAARGEKIGNPFSLTPDDPEP-- 793 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +
	Sbjct: 1048 DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPPIPDYWKPGDSMDAQH 1107
	Query: 794 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 851 L+K P+IVV+VDEFADLMMT GKK+EEELIARLAQKARAAGIHL+LATQRPSVDVITGL
55	Sbjct: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167
	Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQRVHGAFASDEEV 911 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV
60	Sbjct: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227
	Query: 912 HRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVLKTRKASISG 971 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG
65	Sbjct: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286
	Query: 972 VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP 1010 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P
	Sbjct: 1287 VQRQFRIGYNRAARIIEQMEAQGIVSEQGHNGNREVLAP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1  ATGATTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGCG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
     101  TGCTCGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTTG
     151  GCATTGGTCG GCTTCTGGGT C.....
//
10     901  .....A TTGCCATCGG TTTGTTTTTA ATTTACCAAA ACGGGCTGAC
      951  CCTGCTTTT GAAGCCGTGG AAGACGGCAA AATCCATTTT TGGCTCGGAC
     1001  TGTCGCCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTTGCGC
     1051  TGCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAGTCT
     1101  GACATTGAAA GGCGGAAAT GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
     51  ALVGFVW... ..
//
     301  ...IAIGLFL IYQNLTLFF EAVEDGKIHF WLGLLPMHII MFVLALILLR
    351  VRSMPSQPFW QAVGKSLTLK GKG*
20

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGCG
     51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
    101  TGCTCGGCCG TGCCGCCGAC GGGCGTGTG CCATCGATGC CGTGTGGCA
    151  TTGGTCGGCT TCTGGGTCAT CGGTATGACG CCGCTTTTGC TGGTGTGAC
    201  CGCATTTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CCGGACAGCG
    251  AAATGTCGGT CTGGCTATCC TGCGGATTGG CATTGAAACA ATGGATACGC
    301  CCGGTGATGC AGTTTGCCGT GCCGTTTGGC GTTTTGGTTG CCGTCATGCA
    351  GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
    401  TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
    451  AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTGCGAA CCTTCGATAC
    501  CGAATCCGGC ATCATGAAA ACCTGTTCTT CCGCGAACAG GACAAAAACG
    551  GCGGCACAA CATCATCTTC GCCAAAGAAG GTAACCTCTC GCTGAACGAC
    601  AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
    651  CGGACGCGCC GACTACAATC AGGTTTCCTT CCAAAAACCT AACCTGATTA
    701  TCAGCACCAC GCCCAAACCT ATCGACCCCG TTTCCACCG CCGTACCATT
    751  CCGACCGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
    801  GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTGCTTG
    851  CCGTGCCGCT TTCCTATTTT AACCCGCGCA GCGGACATAC CTACAATATC
    901  TTGATTGCCA TCGGTTTGTT TTTAATTTAC CAAAACGGGC TGACCCGTGCT
    951  TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
   1001  CTATGCACAT TATCATGTTT GCCGTTGCAC TCATCCTGTT GCGCGTCCGC
   1051  AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGCGAAAA GTCTGACATT
   1101  GAAAGCGGGA AAATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
     51  LVGFVIGMT PLLLVLTAFI SLTLVLTTRYW RDSEMSVWLS CGLALKQWIR
    101  PVMQFAVPFA VLVAVMQLWV IPWAEIERSRE YAEILKQKQE LSLVEAGEFN
    151  SLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLND
    201  NKRTLELRHG YRSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
    251  PTAQLIGSSN PQHQAEIMWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
    301  LIAIGLFLIY QNLTLTLLFEA VEDGKIHFVW GLLPMHIIME AVALILLRVR
    351  SMPSQPFWQA VGKSLTLKGG K*
50

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

```

5      orf101.pep  MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX
      orf101a     MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRX-AIDAVLALVGFVWXXM
              10      20      30      40      50
              //
10     orf101.pep  .....IAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL
      orf101a     LTVSVLLLCLLAVPLSYFNPRSGHTYINILXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL
              280      290      300      310      320      330
15
      orf101.pep  LPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGKX
      orf101a     LPMHIIMFVIAIVLLRVRSMPSQPFWQAVGKSLTLKGGKX
              120      130      140      150
              340      350      360      370

```

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

```

1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
51 CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
101 TGCTCGGCCN TGGCGCCGAC NGGCGTNTCG CCATCGATGC CGTGTGGCA
151 TTGGTCGGCT TCTGGGTCCN NNGNATGACG CCGCTTTTGC TNGTGTGAC
201 CGCATTTATC AGTACGTTGA CCGTGTTGAC CCGCTACTGG CGNGACAGCG
251 AAATGTCGGT CTGGNTATCC TCGGATTGG CATTGAAACA ATGGATACGC
301 CCGGTGATGC AGTTTGCCGT GCCGTTTGCC GTTTTGTTG CCGTCATGCA
351 GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
401 TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGGGTCAAC
451 AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTCGAAA CCTTCGATAC
501 CGAATCCGGC ATCATGAAAA ACCTGTTCC TCGCGAACAG GACAAAAACG
551 GCGGCGACAA CATCATCTTC NCCAAAGAAA GTAACTTCTC GCTGAACGAC
601 AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
35 651 CGGACGCGCC GACTACAATC AGGTTTCCTT CCNAAACTC AACCTGATTA
701 TCAGCACCAC GCCCAAACCT ATCGACCCCG TTCCACACCG CCGTACNATN
751 CCNACNGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC ANGCGGAATT
801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTCTTG
851 CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC
40 901 TTGANTGCCA TCGGTTTGTT TTAATTTTAC CAAAACGGGC TGACCCTGCT
951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
1001 CTATGCACAT CATCATGTTT GTCATCGCAA TCGTACTTCT GCGCTCCGC
1051 AGCATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGGCGGA AAATGA

```

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

```

1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGXAAD XRXAIDAVLA
51 LVGFVWXXMT PLLLVLTAFI STLTVLTRYW RDSEMSVWXS CGLALKQWIR
101 PVMQFAVPFA VLVAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGGFN
151 SLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF XKESNFSLND
50 201 NKRTLELRHG YRYSGTPGRA DYNQVSFXKL NLIISTTPKL IDPVSHRRTX
251 PTAQLIGSSN PQHXAELMWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301 LXAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHII MFVIAIVLLRVR
351 SMPSQPFWQA VGKSLTLKGG K*

```

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

```

55     orf101a.pep  MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRXAIDAVLALVGFVWXXMT 60
      orf101-1     MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT 60
80     orf101a.pep  PLLLVLTAFISTLTVLTRYWRDSEMSVWXS CGLALKQWIRPVMQFAVPFAVLAVMQLWV 120
      orf101-1     PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLAVMQLWV 120

```

	orf101a.pep	IPWAE LRSREYAEILKQKQELSLVEAGGFNSLGRNGRVYFVETFDTESGIMKNLFLREQ	180
	orf101-1	IPWAE LRSREYAEILKQKQELSLVEAGEFNSLGRNGRVYFVETFDTESGIMKNLFLREQ	180
5	orf101a.pep	DKNGGDNIIFXKESNFS LNDNKRTLELRHGYRYS GTPGRADYNQVSFXKLNLII STTPKL	240
	orf101-1	DKNGGDNIIFAKEGNFS LNDNKRTLELRHGYRYS GTPGRADYNQVSFQKLNLII STTPKL	240
10	orf101a.pep	IDPVSHRRTXPTAQLIGSSNPQHAE LMWRI SLTVSVLLLC LLAVPLSYFNPRSGHTYNI	300
	orf101-1	IDPVSHRRTIPTAQLIGSSNPQHAE LMWRI SLTVSVLLLC LLAVPLSYFNPRSGHTYNI	300
15	orf101a.pep	LXAIGLFLIYQNGLTLLFEAVEDGKIHF LGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	360
	orf101-1	LIAIGLFLIYQNGLTLLFEAVEDGKIHF LGLLPMHIIMFAVALILLRVRSMPSQPFWQA	360
	orf101a.pep	VGKSLTLKGGK	371
20	orf101-1	VGKSLTLKGGK	371

Homology with a predicted ORF from *N. gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

gonorrhoeae:

	orf101.pep	MIYQRNLIKELSFTAVGIFV VLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFWV	57
	orf101ng	MIYQRNLIKELSFTAVGIFV VLLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFWVIGM	59
30		//	
	orf101.pep	IAIGLFLIYQNGLTLLFEAVEDGKIHF LGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	333
	orf101ng	SLTVSVLLLC LLAVPLSYFNPRSGHTYNI LIAIGLFLIYQNGLTLLFEAVEDGKIHF LGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	331
35	orf101.pep	LLPMHIIMFV LALILLRVRSMPSQPFWQAVGKSLTLKGGK	373
	orf101ng	LLPMHIIMFV LALILLRVRSMPSQPFWQAVG	362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial amino acid sequence <SEQ ID 504>:

	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA
	51	LVGFVWIGMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	CGLALKQWIR
	101	PVMQFAVPFA	ILIAVMQLWV	IPWAE LRSRE	YAEILKQKQE	LSLVEAGEFN
	151	NLGKRN GRVY	FVETFDTEG	IMKNLFLREQ	DKNGGDNIIF	AKEGNFSLKD
45	201	NKRTLELRHG	YRYS GTPGRA	DYNQVSFQKL	NLIISTTPKL	IDPVSHRRTI
	251	STAQLIGSSN	POHQAELMWR	ISLTVSVLLC	LLAVPLSYF	NPRSGHTYNI
	301	LIAIGLFLIY	QNGLTLLFEA	VEDGKIHF L	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSQPFWQA	VG...			

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1	ATGATTATC	AAAGAACT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
	51	CATTTTCGTC	GTCCTCTTGG	CGGTGTTGGT	GTCCACGCAG	GCGATCAACC
	101	TGCTTGCCG	CGCAGCTGAC	GGGCGTGTG	CCATCGATGC	CGTGTGGCC
	151	TTAGTCGGCT	TCTGGGTCAT	CGGTATGACC	CCGCTTTTGC	TGGTGTGAC
	201	CGCATTATC	AGCAGCTGA	CCGTATTGAC	CCGCTACTGG	CGCGACAGCG
55	251	AAATGTCGGT	CTGGCTATCC	TGCGGATTGG	CGTTGAAACA	GTGGATACGC
	301	CCGTCATGC	AGTTTGCCGT	GCCGTTTGCC	ATCCTGATG	CCGTCATGCA
	351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTGCG	CAGCCGCGAA	TATGCCGAAA
	401	TTTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAAGCCGG	CGAGTTCAAT
	451	AACTTGGGCA	AGCGCAACGG	CAGgggtttaT	TtcgtcgaaA	CCTTTGACAC
60	501	CGaatccgGC	ATCATGAAAA	ACCTGTtcct	GcGCGAACAG	GACAAAAACG
	551	gcggcgacaa	CATCATCTTC	GCaaaaGAag	gtaactTctc	gctgaaggac

-298-

5
10
1101

```

601 AACAAAcgca cgtcgaATT GCGCCACGGC TACCGTTACA GCGGcaccgcC
651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAacTc aacctgATta
701 TCAGCACCAC GCCCAAacTT ATCGaccCCG TTTCCCACCG CCGCACCATT
751 tcgacCGCCC AAcTgATTGG CAGCAGCAAT CCGCAACATC AGGCAGAATT
801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG
851 CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC
901 TTGATTGCCA TCGGTTTGTT TTTAATTTAC CAAAACGGGC TGACCCTGCT
951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGGCTC GGACTGCTGC
1001 CTATGCACAT CATCATGTTT GTCATCGCAA TCGTACTTCT GCGCGTCCGC
1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGgcgGA AAATGA

```

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

15
20

```

1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
51  LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101 PVMQFAVPFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNNGDNIIF AKEGNFSLKD
201 NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
251 STAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFEL GLLPMHIIMF VIAIVLLRVR
351 SMPSQPFWQA VGKSLTLKGG K*

```

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

25
30
35
40
45
50
55
60

```

      10      20      30      40      50      60
orf101-1.pep  MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT
      |||
orf101ng-1    MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT
      10      20      30      40      50      60

      70      80      90      100     110     120
orf101-1.pep  PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV
      |||
orf101ng-1    PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFA ILIAVMQLWV
      70      80      90      100     110     120

      130     140     150     160     170     180
orf101-1.pep  IPWAE LRSREYAEILKQKQELSLVEAGEFN SLGKRNGRVYFVETFDTESGIMKNLFLREQ
      |||
orf101ng-1    IPWAE LRSREYAEILKQKQELSLVEAGEFN SLGKRNGRVYFVETFDTESGIMKNLFLREQ
      130     140     150     160     170     180

      190     200     210     220     230     240
orf101-1.pep  DKNNGDNIIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL
      |||
orf101ng-1    DKNNGDNIIFAKEGNFSLKDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL
      190     200     210     220     230     240

      250     260     270     280     290     300
orf101-1.pep  IDPVSHRRTIPTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI
      |||
orf101ng-1    IDPVSHRRTISTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI
      250     260     270     280     290     300

      310     320     330     340     350     360
orf101-1.pep  LIAIGLFLIYQNGLTLLFEAVEDGKIHFELGLLPMHIIMFAVALILLRVRSMPSQPFWQA
      |||
orf101ng-1    LIAIGLFLIYQNGLTLLFEAVEDGKIHFELGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA
      310     320     330     340     350     360

      370
orf101-1.pep  VGKSLTLKGGKX
      |||
orf101ng-1    VGKSLTLKGGKX
      370

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
     151  ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
     201  TCGTGTCGTC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
    10  251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
      301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
     401  ATTCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPOYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA PVWQDVRVV AGQNDVAATG DAHSPILNNA AANTSNNNTAN
     101  NGTHIPLFAI DTGKLGXVC QQNHLDDYGR ASRHS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTTAKPOYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      pspA    GGG INA+ TLT+ P G+L+ F + G VVI G GLD D DYTRILS ++I+A
      pspA    GGGLINAASVTLTSGVPVLNNGNLTGFVDVSSGKVVIGGKGLDTSADYTRILSRAAEINA 256

25      orf113  PVWQDVRVVAGQNDVAATGDAHSPILXXXXXXXXXXXXXXXXXGTHIPLFAIDTGKLGGMYA 120
      pspA    VWG+DV+VV+G+N + G + P AIDT LGGMYA
      pspA    GVWGKDVKVVS GKNKLD F D G-----SLAKTASAPSSSDSVTPTVAIDTATLGGMYA 307

30      orf113  NKITLISTVEQAGIRNQGWFFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
      pspA    +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A +++ A+ V N
      pspA    DKITLISTDNGAVIRNKGRIFAATGGVTLSDAGKLSNSGSIDAA---EITISAQTVDN 362

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113  GGGFINASCATLTTAKPOYQAGDLSAFKIR 30
      orf113ng SHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQAGDFSGFKIR 224

40      orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWQDVRVVAGQNDVAATGDAHSPILNNA 90
      orf113ng QGNAVIAGHGLDARDTDFTRILVCQQNHLDDYGRTSRHS 263

45      orf113  IDTGKLGXVCQQNHLDDYGRASRHS 135
      orf113ng DFSGFKIRQGNAVIAGHGLDARDTDFTRILVCQQNHLDDYGRTSRHS 263

```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50      1  MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
      51  SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNIGIP

```

```

101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIOGNPWL
151 TRGEARVVVN QINSSHPSQL NGYIEVGGR AEVVIANPAG IAVNGGGFIN
201 ASRATLTTGQ PQYQAGDFSG FKIRQGNVI AGHGLDARDT DFTRILVCQQ
251 NHLDOYGRTS RHS*

```

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10      1 ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
      51 CAACATTTCAT CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTA
101     101 GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT
      151 TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
201     201 ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTGTGAA ACCGATCCAC
      251 GCTTTGCCAA CTACCGTCAA TGTTTGGGTA GTGACTATAT CcGGACAGC
15      301 CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
      351 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
401     401 GTTTAGAcGG TTATCAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
      451 AATGGCGCGA CTGCGGCACG TtcGATGAAT CTCAGCGTTG GCATTGCATT
501     501 AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGTTTGGTAC
20      551 AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGT3CCA
      601 CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
651     651 GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
701     701 CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
751     751 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
25      801 ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
      851 TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
901     901 CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
951     951 TATCACAGGC AAAGAAAAAG GTGTTT..

```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30      1 ..STGHSEQNYT LPREITRNI LGSFAYESHR KALSHHAPSQ GTELPQSNGI
      51 SLPYTSNSFT PLPSSSLYII NPVNKGYLE TDPFRFANYRQ WLGS DYMLDS
101     101 LKLDPNNLHK RLG DGYYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
      151 NGATAARSMN LSVGIALSAE QVAQLTSDIV WL VQKEVKLP DGGTQTVLVP
201     201 QVYVRVKN GD IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
35      251 DNIGGRIHAQ KSAVTATQDI NNIGMLSAE QTL L L NAGNN INSQSTTASS
      301 QNTQGSSTYL DRMAGIYITG KEKGV..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and pspA protein show 50% aa identity in 325aa overlap:

```

40      Orf115: 1 STGHSEQNYTLPREITRNI LGSFAYESHRKALSHHAPSQGTLPQSNGISLPYTSNSFT 60
      pspA: 778 STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVPVVAENGIHPTFT----- 831
      Orf115: 61 PLPSSSLYIINPVNKGYLETDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYYEQR 120
45      pspA: 832 -LPNSSLFIAIPNNKGYLIETDPAFTDYRWLGS GYMLAALQQDPNHIHKRLGDGYYEQR 890
      Orf115: 121 LINEQIAELTGHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
50      pspA: 891 LVNEQIAKLGTGYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIV 950
      Orf115: 181 WL VQKEVKLPDGGTQTVLVPQVYVRVKN GIDIDGKGALLSGSNTQINVSGSLKN-SGTIAG 239
      pspA: 951 WLENETVTLPDGTTQTVLKPVKYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAG 1009
55

```

-301-

Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLNAGXXXXXXXXXXXX 299
 R ALI+N + N+ G + + A DI N G + AE LLL A
 pspA: 1010 REALILNAQNIKNLQGDLOGKNIFAAAGSDITNTGS-IGAENALLLKASNNIESRSETRS 1068

5 Orf115: 300 XXXXXXXXXXXYLDPMAGIYITGKEG 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from
N.gonorrhoeae:

	orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
		: :	
	orf115ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDGTHREQNYTLPEEITRDISLGSFAYESHK	71
15	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVE	81
		:	
	orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVE	131
20	orf115.pep	DPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	191
25	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLQKEVKLPDGGTQTVLVPQ	201
		:	
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLPDGGTQTVLMPQ	251
	orf115.pep	VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
30	orf115ng	VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
	orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
		:	
35	orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
	orf115.pep	EKGV	325
	orf115ng	EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino
 40 acid sequence <SEQ ID 514>:

	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTLPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGSDYMLGS
45	151	LKLDPNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLQKEVKLP	DGGTQTVLMP
	251	QVYVRVKNGG	IDGKGALLSG	SNTQINVSGS	LKNSGTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIAGQ	ISNQSDQGQT
	401	RLQAGRDINL	DTVQTGKYQE	IHFADNHTI	RGSTNEVGSS	IQTGQDVTL
50	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
	501	GNKLVITDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNNGTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDITIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSMDIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVDLAQQ	AIAVAHKAAK	QFDKAKTTAL
55	701	MPWRLPMQVG	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAAATGGTAA	GTTGCACAAC	TACTGGCGTG
	101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAACA	AAATTATACT
60	151	TTGCCGGAGG	AAATCACACG	CGACATTTCA	CTGGGTTTCAT	TTGCCTATGA
	201	ATCGCATAGC	AAAGCATTAA	GCCGTCATGC	GCCCAGCCAA	GGCACTGAGT
	251	TGCCACAAAG	TAACCGGGAT	AATATCCGTA	CTGCCAAAAG	CAACGGTATT

-302-

301 TCGCTACCTT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
 451 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
 5 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
 601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGAAGCCA
 10 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
 801 GTTGTCAAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
 901 GACAATATCG GTGGCGGTAT TCATGCGCAA AAATCAGCGG TTACGCGCAC
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
 15 1001 TGCTCAATGC GGTAAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
 1101 TATCAGAGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAGACA
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
 1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAA
 20 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCGtatTG
 1351 TCAGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
 25 1501 GGTAAATAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGCCTATGC
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
 1651 CAAGCAGGCA ATCATGTTTC CATTTGGTACA ACCCAAATC AAAGCCAAG
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
 30 1751 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAAACCA ATCCCAAAGC
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
 1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCTTG
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
 1951 AACCATTAA ACAGCAAAAC CACCAAACCC TACGAACAAA AAGGCTTAAC
 35 2001 GGTGGCATTC AGTTCGCCCC TTACCGATTT GGCACAACAA GCGATTGCCG
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
 2151 GCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40 1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
 51 LPPEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPREFANYRQ WLGS DYMLGS
 151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
 45 251 QVYVRVKNNG IDGKALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTL LNAGNN INNSQSSKSS
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSQQGT
 401 RLQAGRDINL DTVTQKGYQE IHFDADNHTI RGTNEVGSS IQTKGDTVLL
 451 SGNNLNAAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
 50 501 GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVIDNGTRI
 551 QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQ AIAVAHKAAN KSDKAKTTAL
 701 MPWRLPMQVG RPIKQAKAHK T*

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

20 30 40 50 60 70
 orf115ng-1.p NEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDISLGSFAYESHK
 orf115 STGHSEQNYTLPREITRNLISLGSFAYESHK
 60 10 20 30
 80 90 100 110 120 130
 orf115ng-1.p ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET
 orf115 ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSSLYIINPVNKGKLVET
 40 50 60 70 80

-303-

5 orf115ng-1.p 140 150 160 170 180 190
 orf115 DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLG DGYEQRLINEQIAELTGHRRLDGYQND
 ||||| 90 100 110 120 130 140

10 orf115ng-1.p 200 210 220 230 240 250
 orf115 EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLPDGGTQTVLMPQ
 ||||| 150 160 170 180 190 200

15 orf115ng-1.p 260 270 280 290 300 310
 orf115 VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK
 ||||| 210 220 230 240 250 260

20 orf115ng-1.p 320 330 340 350 360 370
 orf115 SAVTATQDINNIGGILSAEQ TLLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK
 ||||| 270 280 290 300 310 320

25 orf115ng-1.p 380 390 400 410 420 430
 orf115 EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGR DINLDTVQTGKYQEIHFDADNHTIR
 |||||
 EKGV

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

30 gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length
 = 2273
 Score = 604 bits (1541), Expect = e-172
 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

35 Query: 1 LLVQTEKDGLHNEQTFGEKKVSENGKLHNYWRARRKGHDETHREQNYTLPEEITRDIS 60
 L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
 Sbjet: 739 LIVGTFESALDNDLGTGTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796

40 Query: 61 LGSFAYESHKALSRRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I
 Sbjet: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

45 Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLD PNNLHKRLG DGYEQRLINEQIAELT 180
 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLG DGYEQ+L+NEQIA+LT
 Sbjet: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHHKRLG DGYEQKLVEQIAKLT 900

50 Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLP 240
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
 Sbjet: 901 GYRRLDGYTNDDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLP 960

55 Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
 Sbjet: 961 DGTQTVLKPQVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALIILNAQN 1019

60 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQ TLLNAGNNINNQSTAKSSQNAQGSSTY 359
 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
 Sbjet: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

65 Query: 360 LDRMAGIYITGKEKGVLAQAGKDINIIAGQISNQSDQGQTRLQAGR DINLDTVQTGKYQ 419
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
 Sbjet: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSE DGQTVLNAGGD IRSDTTGISR NQ 1138

70 Query: 420 EIHFADNHTIRGSTNEVGSSIQTGKDVTLLSGNNLNAAAEVGSAGKTLAVYAKNDITI 479
 FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +GRLK + A DI +
 Sbjet: 1139 NTIFSDNYVIRKEQNEVGSTIRGNLSLNAKGDIRIRAAEVGSEQGLKLAAGRDIKV 1198

Query: 480 SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
 +G + +DA K+TGRSGG K +T ++ + A S T +GK+++L +G D + G
 Sbjet: 1199 EAGKAHTETEDALKYTGRSGGKIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

Query: 540 SNVISDNGTRIQAGNHVRIGTTQTQSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318

5 Query: 599 QSNEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSEQSK 1378

10 Query: 659 QTYEQKGLTVAFSSPVT 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QVYEQKGVTVVAISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
 51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
 101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
 151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
 201 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT
 301 CAAGCAGGCA ATCATGTTTCG CATTTGGTACA ACCCAAACCTC AAAGCCAAAG
 351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
 401 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
 25 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT
 501 TGTTGCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
 551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
 601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC
 651 GGTGGCATT C AGTTCCGCCG TTACCGATT GGCACAACAA ...

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
 51 GNKLVIDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTQI
 101 QAGNHVRIGT TQTQSQSETY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
 151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPEGNNTIY AQSIDIQAAH
 35 201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and *pspA* protein show 45% aa identity in 224aa overlap:

40 Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVIDDKAQSH 63
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++
 pspA: 1173 DIRIRAAEVGSEQGRKLKLAAGRDIVKVEAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQ 1232

45 Orf117: 64 HETAQSSTFEKGQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSETYHQT 123
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
 pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGSNIIADNHTILSAKNNIVLKAETRSRSAEMNK 1292

50 Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGSGLKGDTTIVAGKHYEQIGSTVSS 182
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AGKHY Q GST+SS
 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352

Orf117: 183 PEGNNTIYAQSIDIQAHNKLNSNTTQTYEQKXLTVAFSSPVT 226
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
 pspA: 1353 PQGDVGISSGKISIDAAQNRYSEQSKQVYEQKGVTVVAISVPVVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFADADNHTIRGSTNEVGSSSIQTKGDVTLTSGNNLNAAAEVGSAGKTLAVYAKNDITIS	480
10	orf117.pep	AGINTTHVDDASKHTGRSGGGNKLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
	orf117ng	SGIHAGQVDDASKHTGRSGGGNKLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
15	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSSEYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	NVISDNGTRIAGNHVRIGTTQTQSQSSEYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	600
20	orf117.pep	NEHTGSTVGS�KGDTTIVAGKHYEQIGSTVSSPEGNNIYAQSIDIAAHNKLNSNTTQT	210
	orf117ng	NEHTGSTVGS�KGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTTQT	660
25	orf117.pep	YEQKXLTVAFSSPVDLAQQ	230
	orf117ng	YEQKGLTVAFSSPVDLAQQAIIVAHKAAKQFDKAKTTALMPWRLPMQVGRFLKQAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
	51	LPEEITRDIS LGSFAYESH KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
30	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
	151	LKLDPNLHK RLGDGYEQR LINEQIAELT GHRRLDGYON DEEQFKALMD
35	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVOKEVKLP DGGTQTVLMP
	251	QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
40	301	DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGNN INNQSTAKSS
	351	QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSDQGGT
45	401	RLQAGRDLNL DTVQTGKYQE IHFADADNHTI RGSTNEVGSS IQTKGDVTL
	451	SGNNLNAAAEVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
50	501	GNKLVIDKA QSHHETAQSS TFEKQVVLQ AGNDANILGS NVISDNGTRI
	551	QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
55	601	NEHTGSTVGS LKGGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
	651	NQLNSKTTQT YEQKGLTVAF SSPVDLAQQ AIAVAHKAAC QFDKAKTTAL
60	701	MPWRLPMQVG RLFKQAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

45	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGACAAAC TACTGGCGTG
50	101	CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAAAC AAATTATACT
	151	TTGCCGGAGG AAATCACACG CGACATTTCAT CTGGGTTCAT TTGCCTATGA
55	201	ATCGCATAGC AAAGCATTAA GCCGTCATGC GCCAGCCAA GGCAGTGA
	251	TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT
60	301	TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
	351	ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGTA ACCGATCCAC
65	401	GCTTTGCCAA CTACCGTCAA TGTTGGGTA GTGACTATAT GCTGGGCAGC
	451	CTCAAAC TAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
70	501	CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGCATCGTC
	551	GTTTAGACGG TTATCAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT
75	601	AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT
	651	AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGTTGGGTAC
80	701	AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
85	801	GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
	851	CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
90	901	GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
	951	ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
95	1001	TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
	1051	CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA

5
10
15
20

```

1101 TATCACAGGC AAAGAAAAAG GTGTTTGTAGC AGCGCAGGCA GGCAAAGACA
1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAG
1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCctatTG
1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGCGCGC
1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATGTCAG GCAGGAAACG
1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG
1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCTG
1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
2001 GGTGGCATTC AGTTCGCCCG TTACCGATTT GGCACAACAA GCGATTGCCG
2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
2101 ATGCCATGCG GGTCGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
2151 GCGCACAAA ACTTAG

```

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

25
30
35

```

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
51 LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPFRFANYRQ WLGS DYMLGS
151 LKLDPNNLHK RLG DGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
251 QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTL LNAGNN INNQSTAKSS
351 QNAQGSSTYL DRMAGIYITG KEGVLAQA GKDINIIAGQ ISNQSDQGGT
401 RLQAGR DINL DTVQTKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL
451 SGNNLNAAK EVGSAKGLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
501 GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
551 QAGNHVRIGT TQTSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS
601 NEHTGSTVGS LKGD TTIVAS KHYEQTGSNV SSPEGNNLIS TQSM DIGAAQ
651 NQLNSKTTQT YEQKGLT VAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
701 MPWRLPMQVG RPIKQAKAHK T*

```

40 ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

45
50
55
60
65

```

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length = 2273
Score = 604 bits (1541), Expect = e-172
Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHD E TGHREQNYTLPEEITRDIS 60
L+V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I
Sbjct: 739 LIVGTPESALDND ETLGTKTI-TDKGDLHRYHRHKKGRDSTGYSRSPYEPAPEVS-SIR 796

Query: 61 LGSFAYESHKALSRHAPSQGT ELPQSNRD NIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
+G AY+ + AP Q +++P + + NGI +T LP SSL+ I
Sbjct: 797 MGISAYKGY-----APQQASDIPGT V---VPVVAENGIHPTFT-----LPNSSLF AI 840

Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELT 180
P NKGYL+ETDP F +YR+WLGS YML + L+ DPN++HKRLGDGYEQL+NEQIA+LT
Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSYMLAALQQDPNHIHKRLGDGYEQLKLVNEQIAKLT 900

Query: 181 GHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVQKEVKLP 240
G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTL P 960

Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGR NALIINTDT 299
DG TQTVL P+VYVR + ++G+ALLSGS I SG+++N G IAGR ALI+N
Sbjct: 961 DGTQTVLKP KPVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTL LNAGNNINNQSTAKSSQNAQGSSTY 359

```


-307-

+ N+ G + + A D I N G I A E L L L A N N I ++S +S+QN QGS
 Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078
 Query: 360 LDRMAGIYITGKEKGVLAQAQKDINIIAGQISNQSDQGQTRLOAGRDINLDTVQTGKYQ 419
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLISGNNLNAKAAEVGSAKGTILAVYAKNDITI 479
 FD+DN+ IR NEVGSI+I+T+G+++L + ++ +AAEVGS +G L + A DI +
 Sbjct: 1139 NTIFSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198
 Query: 480 SSGIHAGQVDDASKHTGRSGGGNKLVIITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258
 Query: 540 SNVISDNGTRIAGNHVRIGTTTQTSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318
 Query: 599 QSNEHTGSTVGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMIDIGAAQNQLNSKTT 658
 ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378
 Query: 659 QTYEQKGLTVAFFSSPVT 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QVYEQKGVTVAI SVPV 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAAGTTCG
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GTCAATCATG ATGCCGAAAC CCCAACCAGG
 35 201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGCATGCGC AACCTGCAAG
 251 AACAGGATGC CGTCTACATC GCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGAA GAAAGCGGCA TTATCGGCAA
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACGAAAC
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAC ACCTGCAAAA
 40 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
 501 CGTGCGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR
 51 DGKPSGG SVM MPKPQPAVKK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP
 45 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK
 151 PLITLKELSK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAAGTTCG
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GTCAATCATG ATGCCGAAAC CCCAACCAGG
 201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGCATGCGC AACCTGCAAG
 251 AACAGGATGC CGTCTACATC GCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGAA GAAAGCGGCA TTATCGGCAA
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACGAAAC
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAC ACCTGCAAAA
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA
 501 CGTGCGCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
 551 TGCACGCACT GCGCGCCTT TCCAACCGCT GCGCTACCA GATTGTGCGC
 60 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG

-308-

5
10
15
20
25
30
35
40
45
50
55
60

651 CTATCAGGCA TTTATCGTGG GTATTTCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTTCGCA
751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
851 CCATCCATTG GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TCGCCTGTT CTCCTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15
20
25
30
35
40
45
50
55
60

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSNHR
51 DGKPSGGGSM MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK
151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDDHFQIA EPIPIRQYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGFVLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFES
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVLARQSE MLKVGIEPGG KTAALRIFS*

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N.meningitidis*:

30
35
40
45

orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGKPSGGGSM
orf119a	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSNHRDGKPSGGGPM
orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
orf119a	MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY
orf119a	TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE
orf119a	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPIRQYQAFIVGIQAVSRNGLASQEELS

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50
55
60

1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA
51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
101 GGCACCTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCGC CCATGTCGCG
151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCAACCGGCG
201 GGTCAAAAAA ACGGCAAAAT CCAAGACCC CGCCATGCGC AACCTGCAAG
251 AGCAGGATGC CGTCTACATC GCAAGCAGA AACAGGCAA AGCCTCCCCG
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
351 CTCCGCCAC ACCGTTCCCG AACCCCAAAC CGGACATTCC GCACCAAAC
401 CTGCCGACGC GCCGGCAAAA CCTGTTCCCG TTCCGCAAAC GCCGGCAAAA
451 CCGCTGATTA CGCTCAAAGA GCTGTGAAG GTCGAGCTGC CCTGGTTTGA
501 CGTGCCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAGAAC
551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCGGC
601 TGCACCATGG ACGACATTT CCAGATTGCC GAACCCATCC CGGGCATCCG

-309-

5
10
15
20
25
30
35
40
45
50
55
60
65

```
651 CTATCAGGCA TTTATCGTGG GTATTGAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTGCGA
751 CACAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGCGCCTGTT CTCCTAA
```

This encodes a protein having amino acid sequence <SEQ ID 528>:

15
20
25
30
35
40
45
50
55
60
65

```
1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEODAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGVFVLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVRLARQSE MLKVGIEPGG KTALRLFS*
```

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25
30
35
40
45
50
55
60
65

```
orfl19a.pep      10      20      30      40      50      60
MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGGPVM...
orfl19-1          10      20      30      40      50      60
MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGGSV...

orfl19a.pep      70      80      90     100     110     120
MPKPQPAVKKTAKSQDPAMRNLEQODAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
orfl19-1          70      80      90     100     110     120
MPKPQPAVKKTAKPQDPAMRNLEQODAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH

orfl19a.pep     130     140     150     160     170     180
TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWDFVRFDFISYIALTE
orfl19-1         130     140     150     160     170     180
TVSEPTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWDFVRFDFISYIALTE

orfl19a.pep     190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
orfl19-1         190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS

orfl19a.pep     250     260     270     280     290     300
AFNRQVDAFAHSMGGQTLHTDLAAAFIEVASALDAFCARVDQTIAIHLVSP TSISGVELRS
orfl19-1         250     260     270     280     290     300
AFNRQVDAFAQSMGGQTLHTDLAAAFIEVASALDAFCARVDQTIAIHLVSP TSISGVELRS

orfl19a.pep     310     320     330     340     350     360
AVTGVGVFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA
orfl19-1         310     320     330     340     350     360
AVTGVGVFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA

orfl19a.pep     370     380     390     400     410     420
GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG
orfl19-1         370     380     390     400     410     420
GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG
```

```

orfl19a.pep    KTALRLFSX
|||||
orfl19-1       KTALRLFSX

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

```

10 orfl19.pep    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRD GKPSGGG SVM    60
    orfl19ng     MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGG PVM    60

15 orfl19.pep    MPKPQPAVKKKTAKPDPMRNLOEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH    120
    orfl19ng     MPKPQPAVKKPAKPQDSAMRNLOEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH    120

    orfl19.pep    TVSEPQTGHSATKPADASAKPAPVVPQTPAKPLITLKELSKVELSWFDVRIDFISY    175
    orfl19ng     TVSEPQTGHSAPKPADAPAKPVPVQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE    180

```

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

```

20      1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA
      51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
     101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC
     151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCGGC
     201 GGTCAAAAAA CCGGCCAAAC CCCAAGACTC CGCCATGCGC AACCTGCAAG
     25      251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
      301 TTCAAACCG AAATCGAAAC CGCCTTGGAA GAAATCGGCA TTATCGGCAA
     351 CTCGGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC
     401 CTGCCGACGC GCCGGCAAAA CCCGTTCCCG TTCCGCAAAC GCCGGCAAAA
     451 CCGCTGATTA CGCTCAAAGA GCTGTGCAAG GTCGAGCTGC CCTGGTTTGA
     50      501 CGTGCGCTtc gACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
      551 TGCACGCACT GCCGCGCCTT tccAACCGCT GCCGCTACCA GATTGTCCGC
     601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG
     651 CTATCAGGCA TTTATCGTGG GTATCCAGGC AGTCAGCCGC AACGGACTTG
     701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGCGGA CGCATTGCGA
     35      751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
      801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
     851 CCATCCATTT GGTTCGCGC ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
     901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCTACT
     951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
     40      1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
      1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGCAAAAAA CCTTCGACGA
     1101 TTTGTTTATG GATTGGCGG TACGCTGTG CCGTCAGTTG AACCTGAATC
     1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTA
     1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
     45      1251 ACCGGGCGGC AAAACCGCCC TGCGCCTGTT TTCATAA

```

This encodes a protein having amino acid sequence <SEQ ID 530>:

```

50      1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
      51 DGKPSGGPVM MPKPQPAVKK PAKPQDSAMR NLQEQDAVYI AKQKQAKASP
     101 FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK
     151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
     201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFRNQADAFa
     251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
     301 AVTGVGVLE DDGAHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
     351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
     55      401 RTYVLARQSE MLKVGIEPGG KTALRLFS*

```

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

```

60      10      20      30      40      50      60
    orfl19ng     MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGG PVM
    orfl19-1     MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGG SVM
      10      20      30      40      50      60

```

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRN	LQEQDAVYIAKQKQAKASPF	KTEIETALEEIGIIGNSAH			
5	orf119-1	MPKPQPAVKKTAKPQDPAMRN	LQEQDAVYIAKQKQAKASPF	KTEIETALEESGIIGNSAH			
		70	80	90	100	110	120
	orf119ng	TVSEPQTGHSAPKPADAPAKP	VPVPQTPAKPLITLKELSKVEL	PWFDVRFDFISYIALTE			
10	orf119-1	TVSEPQTGHSAPKPADAPAKP	PVPVPQTPAKPLITLKELSKVEL	PWFDVRFDFISYIALTE			
		130	140	150	160	170	180
	orf119ng	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPIPGIRYQAFIV	GIVGQAVSRNGLASQEELS			
15	orf119-1	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPIPGIRYQAFIV	GIVGQAVSRNGLASQEELS			
		190	200	210	220	230	240
	orf119ng	AFNRQADAFQSMGGQTLHTDL	AAFIIEVASALDAFCARVDQT	IAIHLVSPTSISGVELRS			
20	orf119-1	AFNRQVDAFAQSMGGQTLHTDL	AAFIIEVASALDAFCARVDQT	IAIHLVSPTSISGVELRS			
		250	260	270	280	290	300
	orf119ng	AVTGVGVFLEDDGAFHYTDT	SGSTMFSICSLNNEPFTNAL	LDNQSYKGFSMLLDIPHSPA			
25	orf119-1	AVTGVGVFLEDDGAFHYTDT	SGSTMFSICSLNNEPFTNAL	LDNQSYKGFSMLLDIPHSPA			
		310	320	330	340	350	360
	orf119ng	GEKTFDDLFLMDLAVRLSGQL	NLNLVNDKMEEVSTQWLKDV	RTYVRLARQSEMLKVGIEP	GGG		
30	orf119-1	GEKTFDDLFLMDLAVRLSGQL	NLNLVNDKMEEVSTQWLKDV	RTYVRLARQSEMLKVGIEP	GGG		
		370	380	390	400	410	420
	orf119ng	KTALRLFSX					
35	orf119-1	KTALRLFSX					
		429					
40	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGGTACGAT	GAAGCTGCTG	ATTTCCTCCA
50	101	TCGCCCTGAT	TTCATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGGCAATA	TTTyGCAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATTT	TGTAACCGAC	TTCCCGATGG	ACATTTCCGC
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTC	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARHGTEFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM
	51	LVSVTERTKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVGLSAAV
60	101	SLVFNHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAAKLNPIDA
	151	LAQD*				

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

```

      1 ATGTCGGTGC AAGCAGTATT GCGGCACAAA ATGCGTTCGC TTCTGACGAT
      51 GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTCCTGTC GTCGCATTGG
      101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
5      151 AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGGCGACA GGCGCAGCGG
      201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
      251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
      301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
      351 TTTTCGACGTG CGCGGACTGA AGCTGGAAC GGGGCGGCTG TTTGACGAAA
10     401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
      451 GACAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
      501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAAACGCTT
      551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGCTGATG
      601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
15     651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
      701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
      751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCT
      801 CATCGCCCTG ATTTTATTGG TAGTCGGCGG CATCGGCGTG ATGAACATCA
      851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
20     901 ATCGGCGCGC GCGCGGGCAA TATTTTGCG CAGTTTTTGA TTGAGGCGGT
      951 GTTAATCTGC GTCATCGGCG GTTTGGTCGG CGTGGGTTTG TCCGCCGCCG
      1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
      1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
      1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGACG
25     1151 CATTGGCACA GGATTGA

```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

```

      1 MSVQAVLAHK MRSLLTMLGI IIGIASVVSF VALNGSQKK ILEDISSIGT
      51 NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTTLT
      101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
30     151 DKLFA DSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
      201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGT E DFFMNSDSI
      251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
      301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLFVN FVTDFPMDIS
      351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

```

Orf134: 2  RHGTEDFFMNSDXIRQIVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEI 61
          RHG +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EI
40     o648: 496 RHGKKDFFTWNMDGVLTVEKTTRTLQLFLTLVAVISLVVGIGVMNIMLVSVTERTREI 555

Orf134: 62  GIRMAIGARRGNIXQQFLIEAXXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAMSVI 121
          GIRMA+GAR ++ QQFLIEA F+ + + S ++++
45     o648: 556 GIRMAVGARASDVLLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALL 615

Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154
          A CST GI FG++PA AA+L+P+DALA++
o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*:

```

      55     orf134.pep          ARHGTEDFFMNSDXIRQIVESTTGTMKLL
          orf134a      GESHTNSITVKIKDNANTQVAEKGLTDLKARHGTEDFFMNSDSIRQIVESTTGTMKLL
                        210      220      230      240      250      260
                        40      50      60      70      80      90

```

-313-

```

5  orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
   orf134a    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
                                270      280      290      300      310      320

10  orf134.pep  LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
   orf134a    LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
                                330      340      350      360      370      380

15  orf134.pep  LAQDX
   orf134a    LAQDX

```

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

```

1  ATGTCGGTGC AAGCAGTATT GGC GCACAAA ATGCGTTCGC TTCTGACGAT
51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCATTGG
101 GCAACGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
20  151 AACACCATCA GCATCTTCCC AGGGCGCGGC TTCGCGGACA GGC GCAGCGG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGCGCG CACGCTGACT
301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
351 TTTGACGTCG CGCGGGCTGA AGCTGGAAAC GGGGCGGCTG TTTGACGAAA
25  401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
30  651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
35  851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
901 ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
951 GTTAATCTGC GTCATCGGCG GTTTGGTCGG CGTGGGTTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTC
1051 CCGATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
40  1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAACTCAAT CCGATAGATG
1151 CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 536>:

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALNGNSQKK ILEDISSIGT
51  NTISIFPGRG FGDRRSGRIG TLTIDDAKII AKQSYVASAT PMTSSGGTLT
45  101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKL FADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

50 ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

```

orf134a.pep  MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGNSQKKILEDISSIGTNTISIFPGRG
orf134-1     MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGNSQKKILEDISSIGTNTISIFPGRG

55  orf134a.pep  FGDRRSGRIGTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134-1     FGDRRSGRIGTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

60  orf134a.pep  RGLKLETGR LFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1     RGLKLETGR LFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

orf134a.pep  ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
65  orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE

```

-314-

orf134a.pep DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
 orf134-1 DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
 orf134a.pep IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
 orf134-1 IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
 orf134a.pep STGIGIAFGFMPANKAAKLNPIDALAQDX
 orf134-1 STGIGIAFGFMPANKAAKLNPIDALAQDX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

orf134.pep ARHGTEDFFMNSDXIRQIVESTTGTMKLL 30
 orf134ng GESHTNSITVKIKDNANTRVAEKGLAELLKARHGTEDFFMNSDSIRQMVESTTGTMKLL 264
 orf134.pep ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG 90
 orf134ng ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIIGG 324
 orf134.pep LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 150
 orf134ng LVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 384
 orf134.pep LAQD 154
 orf134ng LAQD 388

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

1 ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACCAT
 51 GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCGCTGG
 101 GCAACGGTTC GCAGAAAAAA ATCCTCGAAG ACATCAGTTC GATGGGGACG
 151 AACACCATCA GCATCTTCCC CGGGCGCGGC TTCGGCGACA GGCGCAGCGG
 201 CAAAATCAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
 251 GCTACGTTGC CTCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACC
 301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
 351 TTTCGACGTG CGCGGGCTGA AGCTGGAAAC GGGGCGGCTG TTTGATGAGA
 401 ACGATGTGAA AGAAGACGCG CAAGTCGTCG TCATCGACCA AAATGTCAAA
 451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
 501 GAAACGCCCT TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAACGCTT
 551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
 601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
 651 AGACAATGCC AATACCCGGG TTGCCGAAAA AGGGCTGGCC GAGCTGCTCA
 701 AAGCACGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC
 751 AGGCAGATGG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCCTC
 801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGTGTG ATGAACATTA
 851 TGCTGGTGTC CGTTACCGAG CGCACCAAG AAATCGGCAT ACGGATGGCA
 901 ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
 951 GTTAATCTGC ATCATCGGAG GCTTGGTCCG CGTAGGTTTG TCCGCCGCCG
 1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTCG
 1051 GCGGCATCCG TTATCGGGG GTCGCCTGT TCGACCGGAA TCGGCATCGC
 1101 GTTCGGCTTT ATGCTGCCA ATAAGGCAGC CAAACTCAAT CCGATAGATG
 1151 CATTGGCGCA GGATTGA

This encodes a protein having amino acid sequence <SEQ ID 538>:

1 MSVQAVLAHK MRSLLTMLGI IIGIASVVS VALGNGSQKK ILEDISSMGT
 51 NTISIFPGRG FGDRRSGKIK TLTIDDAKII AKQSYVASAT PMTSSGGTIL
 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
 151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
 201 HQITGESHTN SITVKIKDNA NTRVAEKGLA ELLKARHGT DFFMNSDSI
 251 RQMVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
 301 IGARRGNILQ QFLIEAVLIC IIGGLVGVGL SAAVSLVFNH FVTDFFPMDIS

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5  orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSMGNTNISIFPGRG
   orf134-1     MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSMGNTNISIFPGRG

10 orf134ng      FGDRRSQGIKTTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134-1     FGDRRSQGIKTTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

15 orf134ng      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1     RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

20 orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT
   orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGT

25 orf134ng      DFFMNSDSIRQMVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134-1     DFFMNSDSIRQIVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA

   orf134ng      IGARRGNILQQFLIEAVLICIIIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVAC
   orf134-1     IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC

   orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1     STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E.coli* ABC transporter:

```

sp|P75831|YBJZ ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
648
Score = 297 bits (753), Expect = 6e-80
Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

35 Query: 1 MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXXXXXGNGSQKKILEDISSMGNTNISIFPGRG 60
   M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
40 Sbjct: 260 MAWRALAANKMRTLLTMLGIIIGIASVVSIVVVGDAAKQMVLA DIRSIGTNTIDVYPGKD 319

   Query: 61 FGDRRSQGIKTTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
   FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
   Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

45 Query: 121 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179
   G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
   Sbjct: 380 YGMTFSEGNTFNQEQNGRAQVVVLDSNTRRQLFPHKADVVEVILVGNMPARVIGVAEE 439

50 Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239
   ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
   Sbjct: 440 QQSMFGSSKVLRLVWLPYSTMSGRVMGQSWLNSITVRVKEGFDSAEAEQQLTRLLSLRHGK 499

55 Query: 240 EDFFMNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTKEIGIRM 299
   +DFF N D + + VE TT T++ VVGGIGVMNIMLVSVTERT+EIGIRM
   Sbjct: 500 KDFFTWNMDGVLTVEKTRTLQLFLTAVISLVGGIGVMNIMLVSVTERTREIGIRM 559

60 Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAASVIGAVA 359
   A+GAR ++LQQFLIE F+ + + S +++ A
   Sbjct: 560 AVGARASDVLLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619

   Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
   CST GI FG++PA AA+L+P+DALA++
   Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1   ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT
51  GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCCTTCCT GATTTTGAAA GAACGGATT TCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
10  201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGCGCGATGT
251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC
301 GGCTGGCGCG TCGTGTTTTA CCTTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTGTTGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
15  451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA
501 TATGACCGTC GTTTTTCGG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20  1   ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVFSFLILK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSCIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VESALSAAFF LGEELFWQEI LGMCIISAV
201 F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1   ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GGCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAC
101 AATTGCCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTGCGC TCGGGCTGCG CGCCGTATTG CGTCGGGACA mCTCCGCAC
30  201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCTTTTGGC CACTGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCGG TTTACACGCA GGCGGTGCTG CTCCTTGGTT
401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTCCGCGAG CGGTCAGGAA
35  451 ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAAGTGT CTTGGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTCCGT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
40  701 AAGTCGGCGA CAAATTCACG GTGCCTCGC TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGCTGCGCGC ATTTTCTG GCGGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1   MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS
51  TVALGAAAVL RRDxFRTPHW KNHLNRSMVG TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVMSSVWAT
50  201 LTGWHTLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N.*

meningitidis:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
                        |||||
orf135a      STVALGAAAVLRRDTFRTPHWKNHLSRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIF
                        50      60      70      80      90      100

10     orf135.pep      40      50      60      70      80      90
                        LAVFSFLILKERISVYTQAVLLLGFAVVLLNPSFRSQETAALAGLAGGAMSGWAYLK
                        |||||
orf135a      LAVFSFLILKERISVYTQAVLLLGFAVVLLNPSFRSQETAALAGLAGGAMSGWAYLK
                        110     120     130     140     150     160

15     orf135.pep      100     110     120     130     140     150
                        VRELSSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM
                        |||||
orf135a      VRELSSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM
                        170     180     190     200     210     220

20     orf135.pep      160     170     180     190     200
                        TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVFX
                        |||||
25     orf135a      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLAEELFWQEILGMCIIISGILSSIRPTAF
                        230     240     250     260     270     280

orf135a      KQRLQSLFRQRX
                        290     300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51 GCGGCGCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAC
101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC
35 201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCG TTTACACGCA GCGGTGCTG CTCCTTGTT
40 401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTACAGAA
451 ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCAGGC TGGCGCGTCG
40 551 TGTTTACCT TCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG
601 CTGACCGGCT GGCACACCT GTCTTTTCCA TCGGCAGTTT ATCTGTCTGT
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
45 701 AAGTCGGCGA CAAATTCACG GTTGCCCTCG TTTCTATAT GACCGCTGTT
751 TTTCCGCTC TGTCTGCCG ATTTTTTCTG GCCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
51 TVALGAAAVL RRDFTFRPHW KNHLSRSMVG TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSQE
151 TAALAGLAGG AMSGWAYLKV RELSSLAGEPG WRVVFYLSVT GVAMSSVWAT
55 201 LTGWHTLSFP SAVYLSLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60     orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
                        |||||
orf135-1      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

```

```

5  orf135a.pep  RRDTFRTPHWKHNLRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
   orf135-1    RRDXFRTPHWKHNLRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE

   orf135a.pep  RISVYTQAVLLLGAGVVLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
   orf135-1    RISVYTQAVLLLGAGVVLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG

10  orf135a.pep  WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSMTRAYKVGDKFT
   orf135-1    WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSMTRAYKVGDKFT

15  orf135a.pep  VASLSYMTVVFSAALAAFFLAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
   orf135-1    VASLSYMTVVFSAALAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20 *N.gonorrhoeae*:

```

   orf135.pep  GTGAMLLLFYAVTXLPLATGVTLSTSSIF 30
   orf135ng    STVTLGAAVLRDTRTPHWKHNLRSMVGTGAMLLLFYAVTHLPLTGVTLSTSSIF 335

25  orf135.pep  LAVFSFLILKERISVYTQAVLLLGAGVVLNPSFRSGQETAALAGLAGGAMSGWAYLK 90
   orf135ng    LAVFSFLILKERISVYTQAVLLLGAGVVLNPSFRSGQEPALAGLAGGAMSGWAYLK 395

   orf135.pep  VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM 150
   orf135ng    VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM 455

   orf135.pep  TRAYKVGDKFTVASLSYMTVVFSAALAAFFLGEELFWQEILGMCIIISAVF 201
   orf135ng    TRAYKVGDKFTVASLSYMTVVFSAALAAFFLGEELFWQEILGMCIIISAAF 506

```

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

```

40  1  MPSEKAFRRH LRTASFQGLH LHHFHQKV GK  CGIIGFGIHI FPTLLPAAQC
   51  ILDIQLGLFR IDFAALAVYR RTQVDFIHTV IDGIASDQAF SEVVQILRRL
  101  NLGHFTDTHL IAQARRFIAD FGNIRPMRRG EAKTFCRCFR FDGIDGIHGD
  151  FRQCGHINRL APGKDCRNGK RDKVFFHTRH YNOVCLEKTN CSARKIKFRH
  201  OKQAKTHSTS LAARFTIRPS LSQRPFMDTA KKDILGSGWM LVAAACFTVM
  251  NVLIKEASAK FALGSGELVF WRMLFSTVTL GAAAVLRDRT FRTPHWKHNH
  301  NRSMVGTGAM LLLFYAVTHL PLTTGVTLST TSSIFLAVFS FLILKERISV
45  351  YTOAVLLLG AGVVLNPS FRSGQEPAL AGLAGGAMSG WAYLKVRELS
   401  LAGEPGWRVV FYLSATGVAM SSVWATLTGW HTLSFPSAVY LSGIGVSALI
   451  AQLSMTRAYK VGDKFTVASL SYMTVVFSA SAAFFLGEEL FWQEILGMC
   501  IISAAF*

```

Further work revealed the following gonococcal sequence <SEQ ID 547>:

```

50  1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCC
   51  GGCGGCCTGC TTCACCGTTA TGAACGTATT GATTAAGAG GCATCGGCAA
  101  AATTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTC
  151  ACCGTTACGC TCGGTGCTGC CGCCGTATTG CGGCGCGACA CCTCCGCAC
  201  GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
55  251  TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTGAC AACGGCGTT
   301  ACCCTGAGTT ACACCTCGTC GATTTTTtg GCGGTATTTT CCTTCCTGAT
   351  TTTGAAAGAA CGGATTCCG TTTACACGCA GCGGTGCTG CTCTTGTT
   401  TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTCCGCAG CGGTCAGGAA
   451  CCGGCGGCAC TCGCGGGCT GCGGGCGGC GCGATGTCCG GCTGGGCGTA
60  501  TTTGAAAGTG CGGCAACTGT CTTTGGCGGG CGAACCGGC TGGCGCGTCG
   551  TGTTTTACCT TTCCGCAACC GCGGTGGCGA TGTCGTCggt ttgggcgacg
   601  Ctgaccggct ggCACAccT GTCCTTTcca tcggcagttt ATCtgtCGGG

```

-319-

5
651 CATCGGCGTG tccgcgCtgA TTGCCCAaCT GtcgatgAcg cGCGcctaca
701 aaGTGCGCGA CAAATTCACG GTTGCTCGC tttcctaTAt gaccgtcGTC
751 TTTTCCGCCC TGTCTGCCGC ATTTTCTCTg ggcgaagagc tttTctggCA
801 GGAAATACTC GGTATGTGCA TCATTatcct CAGCGGCATT TTGAGCAGCA
851 TCCGCCCCAT TGCCTTCAAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA
901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

10
1 MDTAKKDILG SGWMLVAAAC FTVMNVLIKE ASAKFALGSG ELVFWRMFLFS
51 TVTLGAAAVL RRDTERTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSGQE
151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT
201 LTGWHTLSEF SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEIL GMCIILSGI LSSIRPIAFK QRLQALFRQR
301 *

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

20
orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLIKEASAKFALGSGELVFWRMFLFSTVTLGAAAVL
orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLFSTVALGAAAVL
20
orf135ng-1.pep RRDTERTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSYTSSIFLAVFSFLILKE
orf135-1 RRDXFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
25
orf135ng-1.pep RISVYTQAVLLLGAFAGVVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG
orf135-1 RISVYTQAVLLLGAFAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
30
orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSEFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT
orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSEFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT
30
orf135ng-1.pep VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIILSGILSSIRPIAFKQRLQALFRQR
orf135-1 VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIILSGILSSIRPTAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40
1 ATGAAGCGGC GTATAGCCGT CTTCGTCCTG TTCCCGCAGA TAATCCGAGT
51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA
101 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT
151 CTGCCCGGGA TCGCCGAAAT CGATTCCCCA TGCGGCATCG TGTTCGGTGC
201 GTCCTCTTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG
45
251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG
301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTCAAGTTCG CCsGGTTCAT
351 TGTTACAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC
401 CACATATGTT CGCAAATTTT GCCGTCTTCG CCGTCTTGA AAAAAGGGAC
451 TTTGACCATG GCAAATCCCA AGGCGGAAAT AATGCGGCGG CGTTCACCAA
50
501 AAAGcTCGCG CCAAAAATAT TTGAATGTTT TACGGGCGCG TTCGTCGGCA
551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC
601 CATCATATCT GTCCTCAAC GTGTACGGTA TCTGTTTGCA CCTTACTGCG
651 GCTTTCTgCc KTCGGCATCC GATTCCGATT TGAAAAGTTC mmrwyATTTCG
701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAVFVL FPQIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY
51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVVR

-320-

101 NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD
 151 FDHGKIQGGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA
 201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE*

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

5 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG
 51 AGTTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
 101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
 151 TATCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
 201 TGCGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAAGCCG
 10 251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCTCAAC
 301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT
 351 CATTGTTTTCAG CACACCGTAA ATATAAAGAC CGTCAAAATA AATATCGTCG
 401 ATCCACATAT GTTCGCAAT TCCGCGTCT TCGCCGCTCTT GGAAAAAAGG
 451 GACTTTGACC ATGGCAAAAT CCAAGCGCGA AATAATGCGG CGGCGTTCCC
 15 501 AAAAAAGCTC GCGCCAAAAA TATTTGAATG TTTTACGGGC GCGTTCGTCG
 551 GCACGGTTTA CCGGTTTCGTC TGCTGTCTT ACATAATAAA TGACGGAATC
 601 GCCCATCATT CTGCTCCTCA ACGGTACGG TATCTGTTTG CACCTTACTG
 651 CGGCTTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT
 701 CGGAATAG

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

1 MMKRRIAEV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ
 51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAVGDAVA HEHPVADVNN
 101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
 151 DFDHGKIQGG NNAAFPKKL APKIFECFTG AFVGTVYRFV CLFYIINDGI
 25 201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

30		10	20	30	40	50	59
	orf136.pep	MKRRIAEVFLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS					
	orf136a	MMKRRIAEVFLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGIAEIDS					
35		10	20	30	40	50	60
	orf136.pep	60	70	80	90	100	119
		PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNNRNANAFALFDIGQFAXFIVQ					
	orf136a	PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVNNRNANAFALFDIGQFAGFIVQ					
40		70	80	90	100	110	120
	orf136.pep	120	130	140	150	160	179
		HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG					
45	orf136a	HAINVKTVKINIVDPHMFANFAVFAVLEKRALTMASKXXXMRRRSQKSSRQKYLNLVLR					
		130	140	150	160	170	180
50	orf136.pep	180	190	200	210	220	230
		AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKSSXXSEX					
	orf136a	R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX					
		190	200	210	220	230	

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
 51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
 101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
 151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
 201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAAGCCG
 251 CCGTAGGGAA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCTCAAC

-321-

5
301 CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT
351 CATTGTTCAG CACGCCATAA ATGTAAAGAC CGTCAAAATA AATATCGTCG
401 ATCCACATAT GTTCGCAAAT TTCGCCNTCT TCGCCGTCTT GGAAAAAAGG
451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NNGATGCGGC GGC GTTCCCA
501 AAAAAGCTCG CGCCAAAAT ATTTGAATGT TTTGCGGGCG CGTTCGCCGG
551 CACGGTTTAC CGGTTTGTCT GCCTGTCTA CATAATAAAT GACGGAATCG
601 CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG
651 CGGCTTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT
701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFOXF GMFFFFIHHQ
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAGNAVA HEHPVADVNN
101 RNANAFALFD IGQFAGFIVQ HAINVKT VKI NIVDPHMEAN FAXFAVLEKR
151 ALTMAKSKXX XMRRRSQKSS RQKYLNVLRA RSPARFTGLS ACST**MTES
15 201 PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

		10	20	30	40	50	60
orfl36a.pep		MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFOXF	GMFFFFIHHQ	QYLP	PGIAEIDS		
20 orfl36-1		MMKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFOIFGM	FFFFIHHQ	QYLP	PGIAEIDS		
		10	20	30	40	50	60
		70	80	90	100	110	120
orfl36a.pep		PCGIVFGTLLFRHXSTHCLY	GKAAGNAVAHEHPVADVNNRNANAFALFDIGQFAGFIVQ				
25 orfl36-1		PCGIVFGALLFRHLPAHCLY	GKAAGDAVAHEHPVADVNNRNANAFALFDIGQFAGFIVQ				
		70	80	90	100	110	120
		130	140	150	160	170	180
orfl36a.pep		HAINVKT VKI NIVDPHMEANFAXFAVLEKRAL	TMAKSKXXXMRRRSQKSSRQKYLNVLRA				
30 orfl36-1		HTVNIKT VKI NIVDPHMEANFAVFAVLEKRDFDHGKI	QGGNNAAFPPKKLAPKIFECFTG				
		130	140	150	160	170	180
		190	200	210	220	230	
orfl36a.pep		R---SPARFTGLSACSTXXMTESPIISAPQVRV	YLFAPYCGFLPSASDSLKS	SKYSE			
35 orfl36-1		AFVGTVYRFVCLFYIINDGIAHH---SAPQVRV	YLFAPYCGFLPSASDSLKS	SKYSE			
		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45	orfl36.pep	MMKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFOIFGM	FFFFIHHQ	QYLP	PGIAEIDS	59
	orfl36ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFOIFGM	FFFFIHHQ	QYLP	PGIAEIDS	60
	orfl36.pep	PCGIVFGALLFRHLPAHCLY	GKAAGDAVAHEHPVADVNNRNANAFALFDIGQFAXFIVQ			119
50	orfl36ng	PGGIVFGTLLFRHLSAHCLY	GKAAGDAVAHEHPVADVNNRNANAFALFDIGQSAGFIVQ			120
	orfl36.pep	HTVNIKT VKI NIVDPHMEANFAVFAVLEKRDFDHGKI	QGGNNAAFPPKKLAPKIFECFTG			179
55	orfl36ng	HTVNIKT VKI NIVDPHMEANFAVFAVLEKRDFDHGKI	QGGNNAAFPPKKLAPKVFEFTG			180
	orfl36.pep	AFVGTVYRFVCLFYIINDGIAHHSAPQVRV	YLFAPYCGFLPSASDSLKS	SKYSE		234
	orfl36ng	AFAGTVYRFVCLFYIINDGIAHHTAPQVRV	YLFAPYRGFLPPASDSLKS	SKYSE		235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60
1 ATGATGAAGC GCGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC

5

10

15

1	MMKRRIVAFV	LLMQKIRILG	QLLPKIVNTV	PAHRMLFQIF	GMFFFFIHRQ
51	YLPGIAEIDS	PGGIVFGTLL	FRHLSAHCLY	GKAAVGDAVA	HEHPVADVAN
101	RNANAFALGD	IGQSAGFIVQ	HTVNIKTVKI	NIVDPHMFAN	FAVFAVLEKR
151	DFDHGKIQQG	NNAAAFPKKL	APKVFECTFG	AFAGTVYRFV	CLFYIINDGI
201	AHHTAPORVR	YLFAPYRGEL	PPASDSDLKS	SKYSE*	

25

```

orf136ng      MMKKRIA VFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGLIAEIDS
               |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
orf136-1      MMKKRIA VFVLFPOIIRVLGQLLPKIVNTVPAHRMLFOIFGMFFFFITHOYLPGLIAEIDS

```

25

orf136ng PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ
| | | | | : | | | | : | | | | :
orf136-1 PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGOFAGFIVO

30

```

orf136ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKI QGGNNAAAF PKKLAPKVFE CFTG
              |||||
orf136-1      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIOGGNNAAAF PKKLAPKIFECFTG

```

35

```
orf136ng      AFAGTVYRFVCLFYIINDGIAHHTAPQVRVYLFAPYRGFLPPASDSDLKSSKYSEX
               ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf136-1      AFVGTVYRFVCLFYIINDGIAHHSAPQVRVYLFAPYCGFLPSASDSDLKSSKYSEX
```

Example 67

45

```

1   ATGGAATAA  TGTAACGTT  TTCAAAATC  AGACCGCTT  TGCAATCGC
51  CGCCGCCGC  TTGCTTGCC  CC.TGCGGAC  GGCGGAAAT  AATGCTGTCC
101 GCAAGCCGT  GCAAACGCC  AAACCCGCC  CAGTGGTCG  TTTGGCACTC
151 GGTGGCGGC  CATCTAAAG  ATTTGCCAT  GTAGGTATTA  TTAAGGTTTT
201 GAAAGAAAC  GGTATTCTG  TGAAGTGGT  TACCGGCACC  TCCGCAGGTT
251 CGATTGTCG  CAACCTTTT  GCATCGGGTA  TGTCGCCCGA  CCGCCTCGAA
301 TTGGAAGCC  AAATTTTAG  CAAAACCGAT  TTGGTCGATT  TAACCTTGTC
351 CACCAATGG  TTTATCAAG  CGCCAAAGCT  GCAAAATTAC  ATCAACCGCA
401 AACTCCGCG  CATGCAGATT  CAGCAGTTTC  GCATCAAAAT  TGCCGCC...

```

1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRLE
101 LEAEILGKTD LVDLTLSTNG FIKGAKLONY INRKLGRMOI QOFPIKFAA.

BNSDOCID: <WO_9924578A2_1>

-323-

1 ATGGAATA TGGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGGAAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTT
 201 GAAAGAAAC GGTATTCTTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 251 CGATTGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AGGGGAATGC
 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
 601 CCCGTCAGTG CCGCCCGGCG GCAGGGGCGG AATTTCTGTA TTGCCGTCGA
 651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT
 801 CGGCGGATTC GATCAGAAAA AACCGCCAT CCGGTTGGGT GAGGAGGCAG
 851 CACGTGCCGC ATTGCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE
 101 LEAEILGKTD LVDLTSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAQQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAARRQGA NFVIAVDISA RPKNISQGF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKPO VLDLGAUVGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
 301 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orfl37.pep		MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTA	KPAAVVGLALGGGASKGFAH				
35	orfl37a	MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA	KPAAVVGLALGGGASKGFAH				
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orfl37.pep	VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTSTNG					
	orfl37a	VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTSTSG					
		70	80	90	100	110	120
		130	140	149			
45	orfl37.pep	FIKGAQLQNYINRKLGRMQIQFPIKFAA					
	orfl37a	FIKGEKLQNYINRKVGGRRIQFPIKFAAVATDFETGKAVAFNQGNAQAVRASAAIPNV					
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAATA TGGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGGAAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTT
 201 GAAAGAAAC GGTATTCTTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 55 251 CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG TAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC
 60 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG

-324-

5
 601 CCCGTCAGTG CCGCCCGGCG GCANGNNNG NATNTCGTGA TTGCCGTCGA
 651 TATTTCCGCC CGTCCGAGCA AAAACATCAG CCAAGGCTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT
 801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10
 1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE
 101 LEAEILGKTD LVDLTLSTG FIKGEKLQNY INRKVGGRRI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY
 15 301 *

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20
 orf137a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH
 orf137a.pep VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
 25 orf137a.pep FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 orf137a.pep FQPVIIGRHTYVDGGLSQPVVPSAARRXXXXXVIAVDISARPSKNISQGFFSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVVPSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV
 orf137a.pep MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
 orf137-1 MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
 35

Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40 orf137.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH 60
 orf137ng MENMVTFSKIRSFLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH 60
 45 orf137.pep VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSTNG 120
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLVDLTLSTSG 120
 orf137.pep FIKGAKLQNYINRKLGRMQIQFPIKFAA 149
 orf137ng FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55
 1 ATGGAATAA TGGAACGTT TTCAAAAATC AGATCATTTT TGGCAATCGC
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GGCGGGAAAC AATGCCGCC
 101 GCAAGCCGCG GCAAACCGCC AAACCGCCG CAGTGGTCGC TTGGCACTC
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTTT
 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 251 CGATAGTCGG CAGCCTTTTG GCATCGGGTA TGTCGCCCCG CCGCCTCGAA
 301 TTGGAAGCCG AGATTTTAGG TAAAACCGAT TTAGTCGATT TAACCTGTGC
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 60 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCATC AAGGGAATGC

-325-

501 CGGGCAGGCG GTTCGTGCTT CCGCCGCCAT TCCCAATGTG TTCCAGCCAG
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
 601 CCCGTCAGTG CCGCTCGGCG GCAGGGGGCG AATTTTCGTGA TTGCCGTCGA
 651 TATTTCGCA CGTCCGAGCA AAAATGTCGG TCAAGGTTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTTGCA AAACGAGTTG
 751 gggcAGGCGG ATGTGGTTAT CAAACCGCag gtTTTGATT TGGGTGCAGT
 801 CGGCGGATTC GATCAGAAAA AGCGCGCCAT CCGGTTGGGC GAGGAGGCAG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSL ASGMSPDRLE
 101 LEAEILGKTD LVDLTLSTG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAQQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV
 15 PVSAAARQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL
 251 GQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY
 301 *

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLAIAAAA LLAACGTAGNNAARKPVQTA KPAAVVALALGGASKGFAH
 orf137-1 MENMVTFSKIRPLLAIAAAA LLAACGTAGNNAVRKPVQTA KPAAVVGLALGGASKGFAH
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSPDRLELEAEILGKTDLVDLTLSTSG
 25 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
 orf137ng FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAQAVRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAQAVRASAAIPNV
 30 orf137ng FQPVIIGRHKYVDGGLSQPVVPSAARRQGANFVIAVDISARPSKNVGQGFSSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVVPSAARRQGANFVIAVDISARPGKNISQGFSSYLDQTLNV
 35 orf137ng MSVSVLQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
 orf137 MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGcTG CCGCTTTCCT
 101 GTCTGCACAC GCTGGGAAAC CCGCTCGGAC ATCTGGCGTT TTACCTTTTA
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CGGGTTTGAA
 201 CCCCAGCCCC AAAACGGTCA AAGCCGTTT TGCGGAAACG GCAAAGGCG
 251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACC GAAGA CATAGAAACA
 50 301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGACAA
 351 ACACGAAGGG CTGCTATTC.

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSCLHTLGN RLGHIAFYLL
 55 51 KEDRARIVAX MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIET
 101 MFKAHVHWEH VQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
   101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
    5    151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGGCAGG  CGGGTTTGAA
       201  CCCCACCCC  AAAACGGTCA  AAGCCGTTTT  TGCAGGAAACG  GCAAAAGGCG
       251  GTTTGGAAGT  TGCCCCGCG  TTTTTCAGAA  AACCAGGAAGA  CATAGAAACA
       301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGACAAA
       351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGCG
   10    401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
       451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
       501  TCGCGGCAAA  GGAATAACCG  CGCCTACCAG  CATACAAGGG  GTCAAAACAA
       551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCGACCAC
       601  GTCCCTCCC  CTCAGGAAGG  CGGGGAAGGC  GTATGGGTGG  ATTTCTTCGG
   15    651  CAAACCTGCC  TATACCATGA  CGCTGGCGGC  AAAATGGGCA  CACGTCAAAG
       701  GCGTGAAAC  CCTGTTTTTC  TGCTGCGAAC  GCCTGCCTGG  CCGACAAGGT
       751  TTCGATTGTC  ACATCCGCCC  CGTCCAAGGG  GAATTGAACG  GCGACAAAGC
       801  CCATGATGCC  GCCGTGTTCA  ACCGCAATGC  CGAATATTGG  ATACGCCGTT
       851  TTCCGACGCA  GTATCTGTTT  ATGTACAACC  GCTACAAAT  GCCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1  MFRLOFRLFP  PLRTAMHILL  TALLKCLSL  LPLSCLHTL  GNRLGHLAF  YLLKEDRA  RIVAX
    51  KEDRARI  VANMRQAG  LNPDPK  TVKAVFA  ETAKGG  LELEPA  AFFRKP  EDIET
   101  MFKAVH  GWEHV  QQALD  KHEG  LLFIT  PHIGS  YDLGG  RYISQ  QLPFL  TAMY
   151  KPPK  IKAID  KIMQAG  RVRG  KGKTA  PTSIQ  GVKQ  IIAL  RSGE  ATIV  LDPH
   201  VPSP  QEGG  EVWVD  FFGK  PAYT  MTLA  AKLA  HVKG  VKTL  FFCC  ERLP  GGQG
   251  FDLH  IRPV  QGEL  NGDK  AHA  AVFN  RNAE  YWIR  RFPT  QYLF  MYNR  YKMP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

      10      20      30      40      50      60
orfl38.pep  MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARI
35 orfl38a    MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARI
      10      20      30      40      50      60
      70      80      90     100     110     120
orfl38.pep  MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
40 orfl38a    MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
      70      80      90     100     110     120
      130     140     150     160     170     180
orfl38.pep  LLF
45 orfl38a    LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTA
      130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
   50  101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
       151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGTCAGG  CAGGCATGAA
       201  TCCCACCCC  AAAACGGTCA  AAGCCGTTTT  TGCAGGAAACG  GCAAAAGGCG
       251  GTTTGGAAGT  TGCCCCGCG  TTTTTCAGAA  AACCAGGAAGA  CATAGAAACA
       301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGACAAA
       351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGCG
       401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
       451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
       501  TCGCGGCAAA  GGAATAACCG  CGCCTACCAG  CATACAAGGG  GTCAAAACAA
       551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCGACCAC
   60

```

5
 601 GTCCCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
 701 GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT
 751 TTCGATTTGC ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC
 801 CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT
 851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

10
 1 MFRLQFRLFP PLRTAMHILL TALLKCLSLL PLSCSLHTLGN RLGLAFYLL
 51 KEDRARIVAN MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIET
 101 MFKAVHGEH VQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY
 151 KPPKIKAIK IMQAGRVRG KGTAPTSIQG VKQIIKALRS GEATIVLPDH
 201 VPSPQEGGEG VWVDFGKPA YTMTLAAKLA HVKGVKTLFF CCERLPGGQG
 251 FDLHIRPVQG ELNGDKAHDA AVFNRNAEYW IRRFPTQYLF MYNRYKMP*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15
 orf138a.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCSLHTLGNRLGLAFYLLKEDRARIVAN
 orf138-1 MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCSLHTLGNRLGLAFYLLKEDRARIVAN
 20
 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG
 orf138-1 MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG
 orf138a.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGTAPTSIQG
 25
 orf138-1 LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGTAPTSIQG
 orf138a.pep VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGVKTLFF
 orf138-1 VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGVKTLFF
 30
 orf138a.pep CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
 orf138-1 CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

40
 orf138.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCSLHTLGNRLGLAFYLLKEDRARIVAX 60
 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCSLHTLGNRLGLAFYLLKEDRARIVAN 60
 orf138.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG 120
 orf138ng MRQAGLNPDPTQTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG 120
 45
 orf138.pep LLF 123
 orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGTAPTGIQG 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50
 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
 55
 201 CCGGACACG CAGACGGTCA AAGCCGTTT TCGCGAAACG GCAAAATGCG
 251 GTTTGGAAC TGGCCCCGCG TTTTTCAAA AACCAGGAAG CATCGAAACA
 301 ATGTTCAAAG CCGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA
 351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGTTG
 401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC
 451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT
 60
 501 GCGCGGCAAA GGCAAAACcg cgcaccg catACAAGGG GTCAAACAAA
 551 tcatcaAGGC CCTGCGCGCG GCGGAGGCAA CCATcATCCT GCGCGACCAC

-328-

5
 601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA
 651 ACCTGCATAc accCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG
 701 TGAAGACCT GTTTTTCTGC TGCAGACGCC TGCCCGACGG ACAAGGCTTC
 751 GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAAGCCCA
 801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTC
 851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10
 1 MFRQLQFRLFP PLRTAMHILL TALLKCLSL SLSCLHTLGN RLGHAFYLL
 51 KEDRARIVAN MRQAGLNPD QTVKAVFAET AKCGLELAPA FFKKPEDIET
 101 MFKAHVHWEH VQQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY
 151 KPPKIKAIK IMQAGRVGK GKTAPTGIQG VKQIIKALRA GEATIILPDH
 201 VPSPQEGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDGQGF
 251 VLHIRPVQGE LNGNKAHDAA VFNRNTEYWI RRFPTQYLFM YNRYKTP*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15
 orf138-1.pep MFRQLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHAFYLLKEDRARIVAN
 orf138ng MFRQLQFRLFPPLRTAMHILLTALLKCLSLLSLSCHTLGNRLGHAFYLLKEDRARIVAN
 20
 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAHVHWEHVQQALDKHEG
 orf138ng MRQAGLNPDQTVKAVFAETAKCGLELAPAFFRKPEDIETMFKAHVHWEHVQQALDKGEG
 25
 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVGKGTAPTISIQQ
 orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVGKGTAPTGIQG
 30
 orf138-1.pep VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGKTLFF
 orf138ng VKQIIKALRAGEATIILPDHVPSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGKTLFF
 35
 orf138-1.pep CCERLPGGQGFDLHIRPVQGE LNKGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP
 orf138ng CCERLPDGQGFVLHIRPVQGE LNKNKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35
 gnl|PID|e334283 (Y14568) htrB [*Pseudomonas fluorescens*] Length = 253
 Score = 80.8 bits (196), Expect = 9e-15
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
 40
 Query: 101 MFKAHVHWEHVQQALDKGEGLLFITPHIGSYD-LGGYISQQLPFHLTAMYKPPKIKAIK 159
 + + V G E +++AL G+G++ IT H+G+++ L Y SQ P Y+PPK+KA+D
 Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVLNHFYCSQCKPI---IFYRPPKLKAVD 150
 Query: 160 KIMQAGRVGKGTAPTGTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGGVWADFFGKPA 219
 ++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A
 45
 Sbjct: 151 ELLRKQRVQLGNKVAASTKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVPFFATQA 208
 Query: 220 YMTLAAKLAHVKGKTLFFCCERLPDGQGF 250
 T + +F RLPDG G+
 Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the
 gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A
 55 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein
 was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1  ..GCGTGGTCGG CCGGCCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
      51  GCATGCGGTG TGGAAACTT TGCCTTCTC GCGGCGGCG GTGTATGCGG
     101  CAGCGGTTT GGGTGTGGTG TATGCGGCGC CGGCGCGGCG GTGCGCGTGG
     151  ATGCGCGGGC TGATGTTTTA GCCGTTTATG GTGTCGCCGG TTTGTGTTTC
     201  GGCGGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTGTC
    10  251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACT GCCGCCGGAT TACGGCAGGG CGGCGGCGGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCAGC TTCCCTCTCT
     401  TGAAACCGGC GTTGCGGCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG
     451  GGCGAATTG CCGCGACATT GTTTCTGTCG CGTCCGGAAT GGCAGACGCT
    15  501  GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
     551  CGCGGGCGAT GGTGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1  ..AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGVV YAAPARRSAW
     51  MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
    20  101  LSAWDALPPD YGRAAGLGA NGFQTACRIT FPLLKPALRR GLTLAAATCV
     151  GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

      1  ATGGATGGAC GCGCTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
     51  GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
    25  101  ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAA
     151  CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
     201  GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GCGCGCGGTG GCGTTTCCGG
     251  GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCCACG
     301  TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
    30  351  GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT
     401  TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGTGCAA
     451  GTGCCTGCGG CACGGCTTCA AACGGCACGG ACGTTGGGCG CGGGGGCGTG
     501  GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
     551  GCGGCGTGTG CCTTGTCTTT CTGTATTGTT TTTCCGGGTT CGGGCTGGCG
    35  601  CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
     651  GTTGGTCATG TTCGAACTCG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
     701  TGGTGTGGG GGTAACGGCG GCGGCAGGGT TGCTGTATGC GTGGTGTGGC
     751  AGGCGCGCGG TTTCCGATAA GCGGTTTCC CCTGTGATGC CGTCGCGGCC
     801  GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTTGCGGCG GCGGTGTGTG
    40  851  CTGTGTGCTG CCTGTTTCCT TTGTTGGCAA TTGTTGTGAA AGCGTGGTGC
     901  GCCGGCGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCGGT
     951  GTGGAATACT TTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT
    1001  TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG
    1051  CTGATGTTTT TGCCGTTTAT GGTGTCGCCG GTTGTGTGTT CGGCGGGCGT
    45  1101  GCTGCTGCTT TATCCGCACT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA
     1151  TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC
     1201  TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTTGGGTGC
     1251  AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCTCTC TTGAAACCGG
     1301  CGTTGCGGCG CCGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
    50  1351  GCGGCGACAT TGTTTCTGTC GCGTCCGGAA TGGCAGACGC TGACGACTTT
     1401  GATTTATGCC TATTTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGGCGA
     1451  TGGTGTGAC ATTGCTGTTG GCGGCGTTTC CGCTGGGTAT TTTCTGCTG
     1501  TTGGACGGCG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA

```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55      1  MDGRRWVWVG AFALLPSAFL AVMVVAPLWA VAAYDGLAWR AVLSDAYMLK
     51  RLAWTVFQAA ATCVLVLPFG VPVAWVLARL AFPGRALVLR LLMLPFVMP
    101  LVAGVGVLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPLV VRAAYQGFVQ
     151  VPAARLQATAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA

```

5

Homology with a predicted ORF from *N.meningitidis* (strain A)

10

15

20

25

30

35

40

45

50

55

60

5	1	MDGRRWAVWG	AFALLPSAFL	AAMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPT
	101	LVAGVGVLAL	FGADGLXWRG	WQDTPYLLY	GNVFFXLPVL	VRAAYQGFVQ
	151	VPAARLQTAX	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLGGSSRYAT	VEVEIYQLVM	FELDMAVASV	LVWLXGVTA	AAGLLYAWFG
10	251	KRAVSDKAVS	VPMPSPQSV	GEYVLLAFAA	AVXSVCCFLX	LLAIVVKAWS
	301	AGESWRVLME	SETWQAVWNT	XRFSAAAVYA	AAVLGVVYAA	AAARSAWMRG
	351	LMFLPFMVSP	VCVSAGVLLL	XPQWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	XDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
	451	AATLFXSRXE	WQTLTLTIYA	YXGRAGXDNY	ARAMVLTLLL	AAFALGXFL
15	501	LDGGEKKRT	ETL*			

	orf139a.pep	MDGRRWVWVGAFALLPSAFLAAMVVPAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA : :
20	orf139-1	MDGRRWVWVGAFALLPSAFLAVMVVPAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
	orf139a.pep	ATCVLVLPPLGVPVAVVLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLALFGADGLXWRG
	orf139-1	ATCVLVLPPLGVPVAVVLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLALFGADGLLWRG
25	orf139a.pep	WQDTPYLLLYGNVFFXLPVLVRAAYQGFGVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP
	orf139-1	RQDTPYLLLYGNVFFNLPVLVRAAYQGFGVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
30	orf139a.pep	WLAGGVCLVFLYCFSGFGLALLGGSRyatVEVEIYQLVMFELDMAVASVLVWLXGVTA
	orf139-1	WLAGGVCLVFLYCFSGFGLALLGGSRyatVEVEIYQLVMFELDMAVASVLVWLVLGVTA
	orf139a.pep	AAGLLYAWFGRRAVSDKAVSPVMPSPPPQSVGEYVLLAFAAAVXSVCCFLXLLAIVVKAWS
35	orf139-1	AAGLLYAWFGRRAVSDKAVSPVMPSPPPQSVGEYVLLAFAAAVLSVCCFLPLLAIVVKAWS
	orf139a.pep	AGESWRVLMESETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMMRGLMFLPFMVSP
	orf139-1	AGESWRVLMESETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMMRGLMFLPFMVSP
40	orf139a.pep	VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLXSAWDALPPDYGRAAGLGANGF
	orf139-1	VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLXSAWDALPPDYGRAAGLGANGF
45	orf139a.pep	QTACRITFPLPKPALRRGLTLAAATCVGEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNY
	orf139-1	QTACRITFPLPKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
50	orf139a.pep	ARAMVLTLLLAALFALGXFLLLDGGEGGKRTETLX
	orf139-1	ARAMVLTLLLAALFALGIFLLLDGGEGGKQETLX

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

	orf139.pep	AWSAGESWRVLMSESETWHA>VWNTLRFSAAA	30
	orf139ng	QSVGEYVLLAFSVAVLSVCCFLPLSAIVVKAWSAGESRRVLMSESETWQAVWNTLRFSAAA	327
60	orf139.pep	VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLLAMYAL	90
	orf139ng	VFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSPVCVSAGVLLLYPGWTASLPLLLAMYAL	387

-332-

```

orf139.pep    LAYPFVAKDVLSAWDALPPDYGAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV    150
|||||
orf139ng      LAYPFVAKDVLSAWDALPPDYGAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV    447

5    orf139.pep    GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL    189
|||||
orf139ng      GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFAVCIFLLLDNGEGG    507

```

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

```

10      1  MDGRCWAVRG  AFSLLPSAFL  AVMVVAPLWA  VAAYDGLAWR  AVLSDAYMLK
      51  RLAWTVFQAA  ATCVLVPLG  VPAVAVLRL  AFPGRALVLR  LLMLPFVMPT
     101  LVAGVGV LAL  FGADG LLWRG  RQDTPYLLLY  GNVFFNLPVL  VRAAYQGFAQ
     151  VPAARLQTAR  TLGAGAWRPF  WDIEMPVLRP  WLAGGVCLVF  LYCFSGFGLA
     201  LLLGGSRYAT  VEVEIYQLVM  FELDMAGASA  LVWLVLGVTA  AAGLLYAWFG
15      251  RRAVSDKAVS  PVMPSPQSV  GEYVLLAFSV  AVLSVCCLFP  LSAIVVKAWS
     301  AGESRRVLME  SETWQAVWNT  LRFSAAVFA  AAVLGVVYAA  AARRLVWMRG
     351  LVFLPFMVSP  VCVSAGVLLL  YPGWTASLPL  LLAMYALLAY  PFVAKDVLSA
     401  WDALPPDYGR  AAAGLGANGF  QTACRITFPL  LKPALRRGLT  LAAATCVGEF
     451  AATLFLSRPE  WQTLTTLIYA  YLGRAGEDNY  ARAMVLTLLL  SAFAVCIFLL
20      501  LDNGEGGKRT  ETL*

```

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

```

      1  ATGGATGGAC  GGTGTTGGGC  GGTACGGGGT  GCTTTTTCCT  TGCTGCCTTC
      51  GGCTTTTTTG  GCGGTAATGG  TCGTTGCGCC  TTTGTGGGCG  GTGGCGGCGT
     101  ATGACGGTGT  GCGGTGGCGC  GCGGTGCTGT  CGGATGCCTA  TATGCTCAAA
25      151  CGTTTGCGCT  GGACGGTGTT  TCAGGCGGCG  GCAACCTGTG  TGCTGGTGCT
     201  GCCTTTGGGC  GTGCCTGTCT  CGTGGGTGCT  GGCGCGGCTG  GCGTTCCTCG
     251  GGCGGGCTTT  GGTGCTGCGC  CTGCTGATGC  TGCCGTTTGT  GATGCCACAG
     301  CTGGTGGCGG  GCGTGGGCGT  GCTGGCTCTG  TTCGGGGCGG  ACGGGCTGTT
     351  GTGGCGCGGC  CGGCAGGATA  CGCCGTATCT  GTTGTGTGAC  GGCAATGTGT
30      401  TTTTCAACCT  GCCCGTGTTG  GTCAGGGCGG  CGTATCAGGG  GTTGTCTCAA
     451  GTGCTGCGCG  CACGGCTTCA  GACGGCACGG  ACGTTGGGCG  CGGGGGCGTG
     501  GCGGCGGTTT  TGGGACATTG  AAATGCCCCG  TTTGCGCCCC  TGGCTTGCCG
     551  GCGGCGGTGT  CCTGTCTTTC  CTGTATTGTT  TTTGCGGGTT  CCGGCTGGCA
35      601  TTGCTGTTGG  GCGGCAGCCG  TTATGCCACG  GTCGAAGTGG  AAATTTACCA
     651  GTTGGTTATG  TTGCAACTCG  ATATGGCGGG  GGCTTCGGCG  CTGGTGTGGC
     701  TGGTGTGGG  GGTAACGGCG  GCGGCAGGGT  TGCTGTATGC  GTGGTTCGGC
     751  AGGCGCGCGG  TTTCGGATAA  GCGGTTTCC  CCCGTGATGC  CGTCGCCGCC
40      801  GCAATCGGTG  GGGGAATATG  TATTGCTGGC  ATTTTCGGTG  GCGGTGTGTT
     851  CCGTGTGCTG  CCTGTTTCCT  TTGTCGGCAA  TTGTTGTGAA  AGCGTGGTCG
     901  GCCGCGCAAT  CGCGGCGGTG  GTTAATGGAA  AGTGAAACGT  GGCAGGCAGT
     951  GTGGAATACT  ttGCGCTTTT  CGGCGGCGGC  GGTGTTTGGC  GCGGCGGCTT
100     1001  TGGGTGTGGT  GTATGCGGCG  GCGGCGCGGC  GGCTGGTGTG  GATGCGCGGA
     1051  CTGGTGTGTT  TACCGTTTAT  GGTGTCGCGC  GTTTGTGTTT  CCGCGGGCGT
45     1101  GCTGCTGCTT  TATCCGGGGT  GGACGGCTTC  GTTACCGCTG  CTGCTGGCGA
     1151  TGTATGCGCT  GCTGCGGTAT  CCGTTTGTGG  CAAAAGATGT  TTTATCGGCC
     1201  TGGGATGCAC  TGCGCGCGGA  TTACGGCAGG  GCGGCGGCAG  GTTTGGCGGC
     1251  AAACGGCTTT  CAGACGGCAT  GCCGTATCAC  GTTCCCCCTC  TTGAAACCGG
     1301  CGTTGCGGCG  CGGTCTGACT  TTGGCGGCGG  CGACGTGTGT  GGGCGAATTT
     1351  GCGGCAACCT  TGTTCCTGTC  GCGTCCGGAA  TGGCAGACGT  TGACGACTTT
50     1401  GATTTATGCC  TATTTGGGGC  GTGCGGGTGA  GGACAATTAT  GCGCGGGCAA
     1451  TGGTGTGAC  ATTGCTGTTG  TCGGCATTTG  CCGTGTGCAT  TTTCTGCTG
     1501  TTGACAACG  GCGAAGGCGg  aaaACGGACG  GAAACGTTAT  AA

```

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

```

55      1  MDGRCWAVRG  AFSLLPSAFL  AVMVVAPLWA  VAAYDGLAWR  AVLSDAYMLK
      51  RLAWTVFQAA  ATCVLVPLG  VPAVAVLRL  AFPGRALVLR  LLMLPFVMPT
     101  LVAGVGV LAL  FGADG LLWRG  RQDTPYLLLY  GNVFFNLPVL  VRAAYQGFAQ
     151  VPAARLQTAR  TLGAGAWRPF  WDIEMPVLRP  WLAGGVCLVF  LYCFSGFGLA
     201  LLLGGSRYAT  VEVEIYQLVM  FELDMAGASA  LVWLVLGVTA  AAGLLYAWFG
60     251  RRAVSDKAVS  PVMPSPQSV  GEYVLLAFSV  AVLSVCCLFP  LSAIVVKAWS
     301  AGESRRVLME  SETWQAVWNT  LRFSAAVFA  AAVLGVVYAA  AARRLVWMRG
     351  LVFLPFMVSP  VCVSAGVLLL  YPGWTASLPL  LLAMYALLAY  PFVAKDVLSA
     401  WDALPPDYGR  AAAGLGANGF  QTACRITFPL  LKPALRRGLT  LAAATCVGEF
     451  AATLFLSRPE  WQTLTTLIYA  YLGRAGEDNY  ARAMVLTLLL  SAFAVCIFLL
     501  LDNGEGGKRT  ETL*

```

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

```

5  orf139ng      MDGRCWAVRGAFSLLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
   orf139-1     MDGRRWVWVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA

10 orf139ng      ATCVLVLPPLGVPAWVVLARLAFPGRALVLRLLMLPFVMPRTLAVAGVGVLAIFGADGLLWRG
   orf139-1     ATCVLVLPPLGVPAWVVLARLAFPGRALVLRLLMLPFVMPRTLAVAGVGVLAIFGADGLLWRG

15 orf139ng      RQDTPYLLLYGNVFFNLPLVLRRAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
   orf139-1     RQDTPYLLLYGNVFFNLPLVLRRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP

20 orf139ng      WLAGGVCLVFLYCFSGFGLALLGGSSRYATVEVEIYQLVMFELDMAGASALVWLVLGVTA
   orf139-1     WLAGGVCLVFLYCFSGFGLALLGGSSRYATVEVEIYQLVMFELDMASVSVLWLVLGVTA

25 orf139ng      AAGLLYAWFGGRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
   orf139-1     AAGLLYAWFGGRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLLAIVVKAWS

30 orf139ng      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
   orf139       AGESWRVLMESETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP

35 orf139ng      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
   orf139-1     VCVSAGVLLLYPQWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF

40 orf139ng      QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
   orf139-1     QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY

50 orf139ng      ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL
   orf139-1     ARAMVLTLLLAFAFGIFLLLDGGEGGKQTETL

```

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

```

45 1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC
   51 GCGGCGCGCA ATCATCCTCA TTCTGATTTT AATCGTCAGA TTCCGCATCC
   101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
   151 GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
   201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
   251 AACGTTTGGT C...

```

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

```

50 1 MDGWTQTLQA QTLGISAIAA IILILILIVR FRIHALLTLV IVSLLTALAT
   51 GLPTGSIVKD ILVKNFGGTL GGVALLVGLG AMLERLV...

```

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

```

55 1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC
   51 GCGGCGCGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
   101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
   151 GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG

```

-334-

201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCCC
 501 GCCCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
 551 GCCAAGTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT
 651 TCCCGAAGT CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
 801 TGCGGACGAA ACCTGGGTTT AGACGGCAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA
 901 CGCGGCGAAA GCGGCGAGCG GTTGGAAAA ACCGTGGACG GCGCACTCGC
 951 CCGCGTCTGT TCCGTGATTG TGATTACCGG CGCGGGCGGT ATGTTCCGGC
 1001 GCGTTTTCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATCGCG
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCCTTGTCT CCTTGGCACT
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACC GCCGCCGCGC
 1151 TGATGGCTCC TGCGGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTCG GTCGGTTGCA GCCACTTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCCTCATCGC ACTCATCGCG
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALTLV IVSLLTALAT
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
 251 IFLNTGVSAL ISEKLVSADL TWVQTAKIIG STPIALLISV LVALFVLGRK
 301 RGESESALEK TVDGALAPVC SVILITGAGG MFGGVLRSAG IGKALADSMA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 401 CIVLATAAGS VGCSEFNDSG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
 451 FALSALLFAI V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N. meningitidis*:

40 orf140.pep MDGWTQTLAQTLLGISAAAIIILILILIVRFRIHALTLVIVSLLTALATGLPTGSIVKD
 orf140a MDGWTQTLAQTLLGISAAAIIILILILIVKFRHALTLVIVSLLTALATGLPTGSIVND
 45 orf140.pep ILVKNFGGTLGGVALLVGLGAMLERLV
 orf140a VLVKNFGGTLGGVALLVGLGAMLRVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
 50

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGACACGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
 51 GCGGCGGCGA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
 101 ACGCGCTGCT GACATGGTTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
 151 GGTTCGCCCA CAGGCAGCAT TGTCACGAC GTACTGGTCA AAAACTTCGG
 201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCCTC
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCCC

-335-

501 GCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTC
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT
 651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
 801 TCGGACGAA ACCTGGGTTT AGACGGCAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGCTT GGGACGCAA
 901 CGCGGCGAAA GCGGCAGCGC GTTGGA AAAA ACCGTGGACG GCGCACTCGC
 951 CCGCTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGC
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
 1051 GATTTGGGCA TTCCGTCCT TTGGGCTGT TTCCTTGTCG CTTGGCACT
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACC GCCGCCGCGC
 1151 TGATGGCTCC TGCCGTGGCC GCGCCCGGCT TTACCGACTG GCAGCTCGCC
 1201 TGATATCGTAT TGGCAACGGC GGCAGGTTTC GTGCGTTGCA GCCACTTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCCGCT CTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCCTCATCGC ACTCATCGGC
 1351 TTTGCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
 25 251 IFLNTGVSAL ISEKLVSAD E TWQTAKIIG STPIALLISV LVALFVLGRK
 301 RGE SGALEK TVD GALAPVC SVILITGAGG MFGGVLRSAG IGKALAD SMA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 401 CIVLATAAGS VGC SHFNDSG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
 451 FALSALLFAI V*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orfl40-1.pep	MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT GLPTGSIVND	60
orfl40a	MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT GLPTGSIVND	60
orfl40-1.pep	ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL IRMFGEKRAP FALGVASLIF	120
orfl40a	ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL IRMFGEKRAP FALGVASLIF	120
orfl40-1.pep	GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP FALASIGAFS VMHVFLPPHP GPIAASEFYG	180
orfl40a	GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP FALASIGAFS VMHVFLPPHP GPIAASEFYG	810
orfl40-1.pep	ANIGQVLILG LPTAFITWYF SGYMLGKVLGR TIHVVPPELL SGGTQDNDL PKEPAKAGTV	240
orfl40a	ANIGQVLILG LPTAFITWYF SGYMLGKVLGR TIHVVPPELL SGGTQDNDL PKEPAKAGTV	240
orfl40-1.pep	VAIMLIPMLL IFLNTGVSAL ISEKLVSAD E TWQTAKIIG STPIALLISV LVALFVLGRK	300
orfl40a	VAIMLIPMLL IFLNTGVSAL ISEKLVSAD E TWQTAKIIG STPIALLISV LVALFVLGRK	300
orfl40-1.pep	RGE SGALEK TVD GALAPVC SVILITGAGG MFGGVLRSAG IGKALAD SMA DLGIPVLLGC	360
orfl40a	RGE SGALEK TVD GALAPVC SVILITGAGG MFGGVLRSAG IGKALAD SMA DLGIPVLLGC	360
orfl40-1.pep	FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA CIVLATAAGS VGC SHFNDSG	420
orfl40a	FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA CIVLATAAGS VGC SHFNDSG	420
orfl40-1.pep	FWLVGRLLDM DVPTTLKTWT VNQTLIALIG FALSALLFAI V	461
orfl40a	FWLVGRLLDM DVPTTLKTWT VNQTLIALIG FALSALLFAI V	461

Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

10	1	<u>MDGRTOTLSA</u>	<u>QTLLGISAAA</u>	<u>IILILILIVK</u>	<u>FRIRALLTLV</u>	<u>IASLLTALAT</u>
	51	<u>GLPTGSIVND</u>	<u>VLVKNFGGTL</u>	<u>GGVALLVGLG</u>	<u>AMLGRLVETS</u>	<u>GGAQSLADAL</u>
	101	<u>IRMFGKEKRAP</u>	<u>FAPGVASLIF</u>	<u>GFPIFFDAGL</u>	<u>IVMLPIVFAT</u>	<u>ARRMKQDVLV</u>
	151	<u>FALASVGAFS</u>	<u>VMHVFLPPHP</u>	<u>GPAASEFYG</u>	<u>ANIGQVLILG</u>	<u>LPTAFITWYF</u>
	201	<u>SGYMLGKVLG</u>	<u>RAIHVPVPEL</u>	<u>LSGGTQSDSD</u>	<u>PKEPAKAGTV</u>	<u>VAVMLIPMLL</u>
15	251	<u>IFLNTGVSAL</u>	<u>ISEKLVSAD</u>	<u>TWVQTAKMIG</u>	<u>STPVALLISV</u>	<u>LAALLVLGRK</u>
	301	<u>RGESGSTLEK</u>	<u>TVDGALAPAC</u>	<u>SVLITGAGG</u>	<u>MFGGVLRASG</u>	<u>IGKALADSMA</u>
	351	<u>DLGIPVLLGC</u>	<u>FLVALALRIA</u>	<u>QGSATVALTT</u>	<u>AAALMAPAVA</u>	<u>AAGFTDWOLA</u>
	401	<u>CIVLATAAGS</u>	<u>VGCSEFNDSD</u>	<u>FWLVGRLSDM</u>	<u>DVPTTLKTWT</u>	<u>VNQTLIAFIG</u>
	451	<u>FALSALLFAI</u>	<u>V*</u>			

	1	ATGGACGGCC	GGACACAGAC	GCTGTCCGCG	CAAACTTGT	TGGGCATTTT
	51	GGCGGCGGCA	ATCATCTCA	TTCTGATTTT	AATCGTCAA	TTCCGCATCC
	101	GCGCGCTGCT	GACACTGGTC	ATCGCCAGCC	TGCTCAGCGG	TTTGGAAC
25	151	GGTTTGCCCA	CAGGCAGCAT	CGTCAACGAC	GTACTGGTCA	AAAACCTTCG
	201	CGGCACGCTC	GGCGGCGTGG	CGCTTCTGGT	CGGTCTGGGC	GCAATGCTCG
	251	GACGTTTGGT	AGAAACATCC	GGCGGCGCAC	AGTCGCTGGC	GGACGCGCTG
	301	ATCGCGGATGT	TCGGCGAAAA	ACGCGCACCG	TTGCTCCGG	CGCTTGCCTC
	351	GCTGATTTTC	GGCTTCCGA	TTTTCTTCTGA	TGCCGGACTA	ATCGCTATGC
30	401	TGCCCATCGT	ATTGCGCCACC	GCACGGCGCA	TGAAACAGGA	CGTACTGCCC
	451	TTGCGCGTTG	CCTCCGTCGG	CGCATTTTCC	GTCATGCACG	TCTTCTGCCC
	501	GCCCCATCCG	GGCCCCGATTG	CCGCTTCCGA	ATTTTACCGG	GCGAACATCG
	551	GCCAGGTTTT	GATTTTGGGT	CTGCCGACCG	CTTCATCAC	ATGTTATTTT
	601	AGCGGCTATA	TGCTCGGCAA	AGTGTGGGG	CGCGCCATCC	ATGTTCCCGT
35	651	TCCCGAAGTG	CTCAGCGGCG	GCACGCAAGA	CAGCGACCCG	CCGAAAGAAC
	701	CTGCCAAGC	AGGAACGGTC	GTGCGCGTCA	TGCTGATTCC	CATGCTGCTG
	751	ATTTTCTTGA	ATACCGCGCT	ATCAGCCCTC	ATCAGCGAAA	AACTCGTAAG
	801	TGCGGACGAA	ACTTGGGTTT	AGACGGCAAA	AATGATCGGT	TCGACACCTG
	851	TCGCCCTTCT	GATTTCCGTA	TTGGCCGCA	TGTTGGTCTT	GGGACGCAAA
40	901	CGCGGCGAAA	GCGGCAGCAT	GTTGGAAAA	ACCGTGAGC	CGCAGCTCGC
	951	CCCCGCCTGT	TCCGTGATT	TGATTACCGG	TCGGGCGGGT	ATGTTGCGCG
	1001	GCGTTTTGCG	CGCTTCCGGC	ATCGGCAAGG	CACTCGCCGA	CAGCATGGCG
	1051	GATTTGGGCA	TTCCCGTCTT	TTTGGGCTGC	TTCTTGTGCG	CCTTGGCACT
	1101	GCGTATCGCG	CAAGGTTCCG	CAACCGTCG	CCTGACCACA	GCCGCGCGCG
45	1151	TGATGGCTCC	CAGGTTGCCC	GCCGCCGGCT	TTACCGACTG	GCAGCTCGCC
	1201	TGTATCGTAT	TGGCAACGGC	GGCAGGTTTG	GTGCGTTGCA	GCCACTTCAA
	1251	CGACTCCGGC	TTCTGGCTGG	TCGGCCGCAT	CTTGATATG	GACGTACCGA
	1301	CCACGCTGAA	AACCTGGAGC	GTCAACCAAA	CCCTCATCGC	ATTCATCGCG
	1351	TTTGCCTTGT	CCGACTGCT	GTTTGCCATC	GTCTGA	

50	1	MDGRTQTL	SA QTL	LGISAAA	IILILILIVK	FRIRALLTLV	IASLLTALAT
	51	GLPTGSIVND	VLVKNFEGTL	GGVALLVGLG	AMLGR	LVETS	GGAQSLADAL
	101	IRMFG	KEKRAP	FAPGVASLIF	GFPIFFDAGL	IVMLPTIVFAT	ARRMKQDVLP
	151	FALASVGA	F	VMHVFLPPHP	GPAAASEFYG	ANIGOVKILG	LPTAFITWYP
55	201	SGYMLGKVLG	RAIHVPVPEL	LSGGTQDSDP	PKEPAKAGTV	VAVMLIPMLL	
	251	IFLNTGVSAL	ISEKLVSAD	TWVQTA	KMIG	STPVALLISV	LAALLVLGRK
	301	RGESGSTLEK	TVDGALAPAC	SVILITGAGG	MFGGYLRASG	IGKALADSMA	
	351	DLGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDWOLA	
	401	CIVLATAAGS	VGCSHFND	SG FWLVGRLLDM	DVPTTLK	TWT VNQTLIAF	IG
	451	FALSALLFAI	V*				

BNSDOCID: <WO __9924578A2_1 >

-337-

```

orfl40-1      MDGWTQTLAQTLGISAAILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
orfl40ng-1.pep VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
5      orfl40-1      ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
orfl40ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASVGAFSVMHVFLPPHPGPIAASEFYG
10      orfl40-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASVGAFSVMHVFLPPHPGPIAASEFYG
orfl40ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPPELLSGGTQDSDPPKEPAKAGTV
15      orfl40-1      ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPPELLSGGTQDNDLPKEPAKAGTV
orfl40ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKMIGSTPVALLISVLAALLVLGRK
20      orfl40-1      VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAIIGSTPIALLISVLAALLVLGRK
orfl40ng-1.pep RGESGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
25      orfl40-1      RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
orfl40ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
30      orfl40-1      FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
orfl40ng-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIAFIGFALSALLFAIV
orfl40-1      FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV

```

30 Furthermore, ORF140ng-1 is homologous to an *E. coli* protein:

```

gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
protein GNTP_BACLI SW: P46832 [Escherichia coli] Length = 454
35      Score = 210 bits (529), Expect = 1e-53
      Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

Query: 88  ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
      E SGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
Sbjct: 80  EHSAGAESLANYFSRKLGDKRTIAALTAAFFLGIPVFFDVGFIILAPIIYGFAKVAKIS 139

Query: 148  VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
      L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K
Sbjct: 140  PLKFGPLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVVGYFAAK 198

Query: 208  VLGRAIHVPVPELL-----SGGTQDSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
      ++ + + E+L G T+ SD P A V ++++IP+ +I T
Sbjct: 199  IINKROYAMSVEVLEQMQLAPASEEGATKLSKINPPGVA-LVTSLIVIPAIIMAGT-- 255

Query: 258  SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
      +S L+ + T ++IGS +RG S + AL
Sbjct: 256  ---VSATLMPPSHPLLGLTQLIGSPMVALMIALVLAFLWLLALRRGWSLQHTSDIMGSALP 312

Query: 318  PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFVALALRIAQGSXXXX 377
      A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
55      Sbjct: 313  TAAVVILVTGAGGVFGKVLVESGVGKALANMLQMLDLPLLPAAFIISLALRASQGS--AT 370

Query: 378  XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFNDSGFWLVGRLLDMDVPTTLK 437
      G Q + LA G +G SH NDSGFV+V + L + V LK
60      Sbjct: 371  VAILTTGGLLSEAVMGLNPICQVLVTLAACFGGLGASHINDSGFWIVTKYLGSLVADGLK 430

Query: 438  TWTVNQTLIAFIGFALSALLFAIV 461
      TWTV T++ F GF ++ ++A++
Sbjct: 431  TWTVLTTILGFTGFLITWCWAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence

65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCGG CGTTCAAACA
      51  TTTGCTGTGC CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
     101  GCGTATTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
     151  AACTTTTTTG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
     201  CTGTATCGGG CTGATTCCAG TTGCCCATT CTCAACCCC GCTGCCGCCG
     251  CCTTTGCCGC CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG
     301  CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
     351  GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCG
     401  TACTGATGTT TTTCCGTCGG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ..DFGISPVYLW VAAAFKHLLS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
      51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAGL VLHGYSLARR
     101  RVIAASFLLG TGWTLSLAA AYPAAFALML PLPVLMEFRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
      51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
     101  TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
     151  GTCGAAGCAC TGGCAGGCAG CCCCACCCCG TTGGTTGCCG ATCTGTTCCG
     201  TCAAACCGAT TCGGCATAC CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
     251  TCAAACATTT GCTGTGCGCG TGGGCTGCCG ACTCATACGA TGCCGCACGC
     301  TTTGCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCCT GCGGCTTTGC
     351  CGGTTTCAAC TTTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTG
     401  TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTTCTT CAACCCCGCT
     451  GCCGCCGCCT TTGCCGCCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
     501  TCGCCGGGCG GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
     551  TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTTGCCCT GATGCTGCCC
     601  TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
     651  GACGGCAGTC GCCTCACTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
     701  CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTCCGCGA ATGGCTCGAC
     751  TATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACGTTG AGACGGCATT
     801  CAGTTTGTGT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
     851  TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
     901  TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTGG TGCTGTTGC
     951  CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCCGCCGC
    1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
    1051  GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
    1101  CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG
    1151  CCGAACGCGC CGCCTATTTT AGCCCGTATT ATGTTCTCTG TATCGATCCC
    1201  ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTGGC TGTGGGCGAT
    1251  TACCCGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
    1301  GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
    1351  GACGCGGCGA AAAGCCACGC GCCGTCGTC CGGAGTATGG AGGCATCGCT
    1401  TTCCCCGGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
    1451  TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
    1501  TTGCCGCACC GCGTCGGCGA TGTACAATGC CGTACCGCA TCGTCTCTCT
    1551  GCCCCAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
    1601  CGCGTCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAATCGGG
    1651  GAAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPPDAR PPAKTHEKPW LLLLMAFAWL WPGVFSHDLW NPDEPAVYTA
      51  VEALAGSPTP LVAHLFGQTD FGIPPVYLVW AAAFKHLLSP WAADSYDAAR
     101  FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPVAHFLNPA
     151  AAAFAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
     201  LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD

```


Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

BNSDOCID: <WO__9924578A2_1_>

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG
 1651 GAAAATATAT TAAAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

```

      1 MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
5      51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLLSP WAADPYDAAR
      101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA
      151 AAFAAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
      201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD
      251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWFA LPALPLAVWT VCRTLRFSTD
10      301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRGAA
      351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP
      401 IPMAVAVLFT PLWLWAIKTR NIRGRQAVTN WAAGVTLTWA LLMTLFLPLWL
      451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT
      501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKGT
15      551 ENILKTTD*
  
```

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

```

      orf141a.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
      orf141-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
20      orf141a.pep LVAHLFGQIDFGIPPVYLWVAAFAKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
      orf141-1 LVAHLFGQIDFGIPPVYLWVAAFAKHLLSPWAADSYDAARFAGVFFAVVGLTSCGFAGFN
      orf141a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPAFAAAGLVLHGYSLARRRVIAASFLLGT
      orf141-1 FLGRHHGRSVVLILIGCIGLIPVAHFLNPAFAAAGLVLHGYSLARRRVIAASFLLGT
25      orf141a.pep GWTLMSLAAAYPAAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT
      orf141-1 GWTLMSLAAAYPAAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT
      orf141a.pep QPALFAQWLDHVFVGTGGVRHIQTAFSLFYLLKNLLWFALPALPLAVWTVCTRLRFSTD
      orf141-1 QPALFAQWLDHVFVGTGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCTRLRFSTD
30      orf141a.pep WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRGAAAFVNWFGIMA
      orf141-1 WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRGAAAFVNWFGIMA
      orf141a.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMNAVAVLFTPLWLWAIKTRK
      orf141-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMNAVAVLFTPLWLWAIKTRK
40      orf141a.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPLWLDAAKSHAPVVRSMELKRELSDGIE
      orf141-1 NIRGRQAVTNWAAGVTLTWALLMTLFLPLWLDAAKSHAPVVRSMELKRELSDGIE
      orf141a.pep CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRIYRIVRLPQNADAPQGWQTVWQGARPRNKD
      orf141-1 CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRIYRIVRLPQNADAPQGWQTVWQGARPRNKD
50      orf141a.pep SKFALIRKTGENI
      orf141-1 SKFALIRKIGENI
55
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

```

60      orf141.pep DFGISPVYLWVAAFAKHLLSPWAADSYDVA 30
      orf141ng WNPAPPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAFAKHLLSPWAADSYDVA 126
  
```

-341-

orf141.pep RFAGVFFAVIGLTS CGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL 90
 orf141lng RFAGVFFAVIGLTS CGFAGFNFLGRHHGRSVVLIHIGCIGLIPVAHFFNPAAAAFAAAGL 186

5 orf141.pep VLHGYSLARRRVIAASFL LGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP 140
 orf141lng VLHGYSLARRRVIAASFL LGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPQSRRLMLTA 246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10 1 MPSEAVSARP LCEYLLHLAI RPFLTLMLT YTPPDARPPA KTHEKPWLLL
 51 LMAFAWLWPG VFSHDLWNPA EPAVYTAVEA LAGSPTPLVA HLFGQTDGFI
 101 PPVYLWVAAA FKHLSPWAA HPYDAARFAG VFFAVIGLTS CGFAGFNFLG
 151 RHHGRSVVLI HIGCIGLIPV AHFFNPAAAA FFAAGLVLHG YSLARRRVIA
 201 ASFLLGTGWT LMSLAAAYPA AFALMLPLPV LMFFRPWQSR RLMLTAVASL
 15 AFALPLMTVY PLLLAKTQPA LFAQWLNHYV FGTFGGVRHI QRAFSLFHYL
 301 KNLLWFAPPG LPLAVWTVCR TRLFSTDWGI LGIVWMLAVL VLLAFNPQRF
 351 QDNLVWLLPP LALFGAAQLD SLRRGAAAFV NWFGIMAFGL FAVFLWTGFF
 401 AMNYGWPAKL AERAAVFSFY YVPDIDIPM AVAVLFTPLW LWAITRKNIR
 451 GRQAVTNWAA GVTTLTALLM TLFLPWLDAA KSHAPVVRSM EASFSPCLKR
 20 501 ELSDGIECIG IGGGDLHTRI VWTQYGTLPV RVGDVRCRYR IVRLPQNADA
 551 PQGWQTVWQG ARPRNKDSKF ALIRKIGENI LKTTD*

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

1 ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
 51 AAAACCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGCTG TGGCCCGCG
 25 101 TGTTTTCCCA CGATTTGTGG AATCCTGCCG AACCTGCCGT CTATACCGCC
 151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCG TTGGTTGCCC ATCTGTTCCG
 201 TCAAACCGAT TTCGGCATA CCCCCGTGTA TCTTTGGGTT GCCCGCCGAT
 251 TCAAACATTT GCTGTCGCCG TGGCAGCCG ACCCGTATGA TGCCGACCG
 301 TTTGCAGCGG TATTTTTTGC CGTTATCGGA CTGACTTCTT GCGGCTTTGC
 30 CGGTTTCAAC TTTTGGGCA GACACCAGG GCGCAGCGTT GTTTTAATCC
 401 ATATCGGCTG TATCGGGCTG ATTCCGGTTG CCCATTTCCT CAATCCcgcc
 451 gccgcccgcct tGCGCGCGC CGGACTGGTG CTGCacgct actcgctgC
 501 ACGCCGCGCG GTGATtgccg cctctTtccT GCTCGGTACG GGTGGACGT
 551 TGATGTCGCT GCGCGCAGCT TATCCGGCGG CGTTTGCGCT GATGCTGCCC
 35 601 CTGCCCCGTC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
 651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGCTTACC
 701 CGCTGCTCtt gGCAAAACG CAGCCCGCGC TGTGTCGCA ATGGCTCAAC
 751 TATCACGTTT TCGGTACGtT cggcgGCGTG CGGCAcaTTC AGAggGCatT
 801 Cagtttgttt cactatctgA Aaaatctgct ttggttcgca ccgcccgggG
 40 851 TGCCGCTGGC GCTTTGGACG GTTTGCCGCA CACGCCTGTT TTCGACCGAC
 901 TGGGGGATTT TGGGCATTGT CTGGATGCTT GCCGTTTTGG TGCTGCTCGC
 951 CTTTAATCCG CAGCGTTTTT AAGACAACCT CGTCTGGCTG CTGCCGCCGC
 1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
 1051 GCTTTTGTCA ACTGTTTCGG CATTATGGCG TTCGGGCTGT TTGCGGTGTT
 45 1101 CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG
 1151 CCGAACGCGC CGCCTACTTC AGCCCGTATT ACGTTCCCGA CATCGATCCC
 1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGC TGTGGGCGAT
 1251 TACCCGGA AAACATACGG GCAGGCAGGC GGTTACCAAC TGGCGGGCAG
 1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTTCT GCCGTGGCTG
 50 1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGTT
 1401 TTCCCGGAA TTAACCGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
 1451 TAGCGGCGCG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
 1501 TTGCCGCACC GCGTCGGCGA TGTCCGTTGC CGCTACCGTA TCGTCCGCCT
 1551 GCCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGGTG
 55 1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTTG CACTGATACG GAAAATCGGG
 1651 GAAAATATAT TAAAAACAAC AGATTGA

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

60 1 MLTYTPPDAR PPAKTHEKPW LLLLMAFAWL WPGVFSHDLW NPAEPAVYTA
 51 VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAFKHLSP WAADPYDAAR
 101 FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLIHIGCIGL IPVAHFLNPA
 151 AAFAAAAGLV LHGYSLARRR VIAASFL LGT GWTLMSLAAA YPAAFALMLP
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLN
 251 YHVFGTFGGV RHQRAFSLF HYLKNLLWFA PPGLPLAVWT VCTRFLSTD
 301 WGILGIVWML AVLVLAFNP QRFQDNLVWL LPPLALFGAA QLDLRRGAA

-342-

351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP
 401 IPMAVAVLET PLWLWAIKTRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
 451 DAAKSHAPVV RSMEASFSPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
 501 LPHRVGDVRC RYRIVRLPQN ADAPOGWQTV WQGAPRNKD SKFALIRKIG
 551 ENILKTTD*

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

```

  orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP
  orf141-1       MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNDEPAVYTAVEALAGSPTP
10
  orf141ng-1.pep LVAHLFGQTDGFI PPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVIGLTSCGFAGFN
  orf141-1       LVAHLFGQTDGFI PPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN
15
  orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLG
  orf141-1       FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLG
20
  orf141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
  orf141-1       GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
25
  orf141ng-1.pep QPALFAQWLNHYVFGTFGGVRHQRQAFSLFHYLKNLLWFAPPGLPLAVWTVCTRLEFSTD
  orf141-1       QPALFAQWLDYHVFGTFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCTRLEFSTD
30
  orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
  orf141-1       WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
35
  orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMVAVVLETPLWLWAIKTRK
  orf141-1       FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMVAVVLETPLWLWAIKTRK
40
  orf141ng-1.pep NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAAKSHAPVVRSMESFSPELKRELSDGIE
  orf141-1       NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAAKSHAPVVRSMESLSPELKRELSDGIE
45
  orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLPHRVGDVRCRYRIVRLPQNADAPQGWQTVWQGAPRNKD
  orf141-1       CIGIGGGDLHTRIVWTQYGTLPHRVGDVQCRIYRIVLLPQNADAPQGWQTVWQGAPRNKD
  orf141ng-1.pep SKFALIRKIGENILKTTDX
  orf141-1       SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

  1  ..CAATCCGCCA AATGGTTATC GGGCCAAACT CTAGTCGGCA CAGCAATTGG
  51  GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
  101 CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
  151 AGCGGTTTTT AGGTAGGCTA TACGTTTTAA

```

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

  1  ...QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
  51  SGFQVGYTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

-343-

1 ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTT TATGTAAATT
 101 ATGGACGTTT GATTGGCGGT ACGCCCGATG AGGAAAGTTT TGACGGCCAT
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT
 5 CCGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT
 301 ACTGATTTTC GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCGGT GTAAACTGT GGATGAGGGA AACAAAAAGT TACATTGATG
 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CTGCGGGTTG GTTGGCAGAA
 10 451 CTTTCCACAA AAGAATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA
 501 ATATAAACGC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTGCCT ATGACACATC
 15 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG
 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGAGGACAT GTTTCAGGAC
 851 AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGTCGGCAC AGCAATTGGG
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTAC
 20 951 CGGCCGCGCA TTGAAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAA
 1001 GCGGTTTTCA GGTAGGCTAT ACGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH
 51 RKEGGSNNYA VHYSAPFGKW TWFNHNHGYR YHQAUSGLSE VYDYNKSYN
 101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE
 151 LSHKEYIGRS TADFKLYKR GTGMKDALRA PEEAFGEGETS RMKIWTASAD
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
 251 SAERGWIWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAIG
 301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF*

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35 orf142.pep QSAKWLSGQTLVGTAIGIRGQIKLGGNLHY 30
 orf142ng RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY 313
 orf142.pep DIFTGRALKKPEFFQSRKWASGFQVGYTF 59
 40 orf142ng DIFTGRALKKPEYFQTKKWTGFQVGYSF 342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC
 51 TTTCTCTGCC GACAATCCTT TTGGACTGAG TGATATGTTT TATGTAAATT
 45 101 ATGGACGTTT AATTGGCGGT ACGCCCGATG AGGAAAATTT TGACGGCCAT
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT
 201 CCGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAC
 301 ACTGATTTTC GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCAGT GTAAACTGT GGACGAGGGA AACAAAAAGT TACATTGATG
 50 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CCACAGGTTG GTTGGCAGAA
 451 CTTTCCACAA AAGGATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA
 501 ATATAAACAC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTGCCT ATGACACATC
 55 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG
 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 CCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGAGGACAT GTTTCAGGAC
 851 AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGCCGGCAC AGCAATTGGG
 60 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTAC
 951 CGGCCGTGCA TTGAAAAAGC CCGAATATTT TCAGACGAAG AAATGGGTAA

1001 CGGGGTTTCA GGTGGGTTAT TCGTTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

```

1 MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH
51 RKEGGSNNYA VHYSAPFGKW TWA FNHNGYR YHQA VSGLSE VYD YNGKSYN
101 TDFGFNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTGWLAE
151 LSHKGYIGRS TADFKLKYKH GTGMKDALRA PEEAFGEGTS RMKIWTASAD
201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
251 PAERGWIWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWL GQTLAGTAIG
301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGFGVGY SF*
```

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

```

15 orf142-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
   orf142ng-1 MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA

   orf142-1.pep VHYSAPFGKWTWAFNHNHNGYRYHQAVSGLSEVYD YNGKSYNTDFGFNRLLYRDAKRKTYLG
   orf142ng-1 VHYSAPFGKWTWAFNHNHNGYRYHQAVSGLSEVYD YNGKSYNTDFGFNRLLYRDAKRKTYLS

20 orf142-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTGMKDALRA
   orf142ng-1 VKLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRA

   orf142-1.pep PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
   orf142ng-1 PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT

25 orf142-1.pep VRGFDGEMSLSAERGWIWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAG
   orf142ng-1 VRGFDGEMSLPAERGWIWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG

   orf142-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF
   orf142ng-1 IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGFGVGYSF
```

In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

```

40 gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
   Score = 119 bits (295), Expect = 3e-26
   Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)

   Query: 2 DMSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
   DMSG ++TG+ Q N + + DN FGL+D ++++ G S + + D + G
   Sbjct: 230 DMSGQKSTGEEQLNGSLALDNVFGGLADQWFISAGHS---SRFATSHDAESLQAG----- 280

45 Query: 62 HYSAPFGKWTWAFNHNHNGYRYHQAVSGLSEVYD YNGKSYNTDFGFNRLLYRDAKRKTYLSV 121
   +S P+G W +N++ RY + G S F +R+++RD KT ++
   Sbjct: 281 -FSMPYGYWNLGYNYSQSRYNRTFINRDFPWHSTGSDTHRFSLSRVVRFDGTMKTAIAG 339

50 Query: 122 KLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRAP 181
   R +Y++ + L RK + ++H + A F Y G +
   Sbjct: 340 TFSQRTGNNYLNGSLLPSSSRKLSSVSLGVNHSQKLWGGLATFNPTYNRGVRWLGETDT 399

   Query: 182 EEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241
   +++ E + WT SA P Y S++ Q++ L ++L +GG ++
   Sbjct: 400 DKSADEPRAEFNKWTLASYYHPV---TDSITYLGSLYGQYSARALYGSEQLTLGGESSI 456

   Query: 242 RGF DGEMSLPAERGWIWRNDLSWQFKP----GHQLYLGA-DVGHVSGQSAKWLSGQTLAG 296
   RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G
   Sbjct: 457 RGF-REQYTSGNRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAASLWG 515

60 Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGFGVGYSF 342
   A+G+ + L + G + P + Q V G++VG SF
```

Sbjct: 516 GAVGMTVASRW---LSQQVTVGWPIISYPAWLQPD TMVVG YRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 73

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 609>:

```

1   ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GgCGGACACC
51  GCCGACATCG ATACCGCTTT GAACCTGTTG TACCGTTTGC AAAAACTCGA
101 ATTCTCTAT  GCGGATGAAA ACGGTCATTC AGACGGCATC AATTGwCGG
151 ACGAGCAATT  GCGTTGCTG ATGGAACAAT TGTCCGGCAG CGGTAAGGCG
201 TTATTGGTCG  ATCGGAACGG TCTGTATCTT GCCAACGCCA ATTTCCATCA
251 TGAGGCGGCG  GAAGAGTTGG GGTGTGTGGC GGCAGAAGTC GCACAGATGG
301 AAAAGAAATA  CCGGCTGCTG ATTAAGAACA AC..

```

This corresponds to the amino acid sequence <SEQ ID 610; ORF143>:

```

15  1   MRTKWSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
51  51  EQLPLLMEQL  SGSGKALLVD RNgLYLANAN FHHEAAEELG LLAAEVAQME
101 101  KKYRLLIKNN ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 611>:

```

20  1   ATGGAATCAA CACTTTCACT ACAAGCAAAT TTATATCCCC GCCTGACTCC
51  51  TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCCGTAAAA
101 101  CTTGTGTGCA CAGCCTGTTG AAAGCAGATG CGGACGAAAT GGTCAGCAGT
151 151  GAGAAGCTGC TTACTTGGGC GGACACCGCC GACATCGATA CCGCTTTGAA
201 201  CCTGTTGTAC CGTTTGCAA AACTCGAATT CCTCTATGGC GATGAAAACG
25  251  GTCATTCAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
301 301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
351 351  GTATCTTGCC AACGCCAATT TCCATCATGA GCGGCGGAA GAGTTGGGGT
401 401  TGTGGCGGC  AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCTGATT
451 451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
501 501  CGGTCAGAGC GAATTGACAT TTTTCCCAT TATATCGGT TCAACCAAT
30  551  TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTTGT
601 601  ACTTTGGTAA GGATTTTATA CCGCCGTTAC AGCAACCGCG TGTA

```

This corresponds to the amino acid sequence <SEQ ID 612; ORF143-1>:

```

35  1   MESTLSLQAN LYPRLTPAGA FYAVSSDAPS AGKTLHLSLL KADADEMVSS
51  51  EKLLTWADTA DIDTALNLLY RLQKLEFLYD DENGHSDGIN LSDEQLPLLM
101 101  EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLLI
151 151  KNNLYINNA  WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV
201 201  TLVRILYRRY SNRV*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of *N.meningitidis*:

```

                                     10      20      30
orfl43.pep      MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFL
                                     | : : | | | | | | | | | | | | | | | |
45  orfl43a      GAFYAVSSDXPSAGKTLHLSLLKADADEMVSSEKLLTWAXTADIDTALNLLYRLQKLEFL
                                     20      30      40      50      60      70

                                     40      50      60      70      80      90
orfl43.pep      YGDENGHSDGINLXDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHEAAEELGLLAAE
                                     | | | | | | | | | | | | | | | | | | | | | |
50

```

-346-

```

orfl43a      YGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHAAEELGLLAAE
              80          90          100          110          120          130

              100          110
5  orfl43.pep  VAQMEKKYRLLIKNN
              |||||
orfl43a      VAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFFPLYIGSTKFILVIGGIPDLGKEA
              140          150          160          170          180          190

```

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

```

10      1  ATGGAATCAA  CANTTTCACT  ACAAGCAAAT  TTATATCNCC  GCCTGACTCC
      51  TGCCGGTGCA  TTTTATGCCG  TATCCAGCGA  TGNCCCCAGT  GCCGGTAAAA
     101  CTTGTTGCA  CAGCCTGTTG  AAAGCGGATG  CGGACGAAAT  GGTNAGCAGT
     151  GAGAAGCTGC  TTACCTGGGC  GGANACCGCC  GACATCGATA  CCGCTTTGAA
     201  CCTGTTGTAC  CGTTTGCAA  AACTCGAATT  CCTCTATGGC  GATGAAAACG
15      251  GTCATTGAGA  CGGCATCAAT  TTGTCGGACG  AGCAATTGCC  GTTGCTGATG
     301  GAACAATTGT  CCGGCAGCGG  TAAGGCGTTA  TTGGTCGATC  GGAACGGTCT
     351  GTATCTTGCC  AACGCCAATT  TCCATCATGA  GGCGGCGGAA  GAGTTGGGGT
     401  TGTGGCGGCG  AGAAGTCGCA  CAGATGGAAA  AGAAATACCG  GCTGCNNATT
     451  AAGAACAACC  TGTATATCAA  CAATAACGCT  TGGGGCGTTT  GCGATCCTTC
20      501  CGGTCAGAGC  GAATTGACAT  TTTTCCCATT  GTATATCGGT  TCAACCAAT
     551  TTATTTTGGT  TATCGGCGGC  ATTCCCGATT  TGGGCAAAGA  GGCATTTGTT
     601  ACTTTGGTAA  GGATNTTATA  CCNCCNGTTA  CAGCAACCGC  GTGTAAAACT
     651  TGGGAGAGAG  GANGGTTTAT  GCAGCAATTA  TTGA

```

This encodes a protein having amino acid sequence <SEQ ID 614>:

```

25      1  MESTXSLQAN  LYXRLTPAGA  FYAVSSDXPS  AGKTLHSL  LKADADEMVS
      51  EKLLTWAXTA  DIDTALNLLY  RLQKLEFLYG  DENGHS DGIN  LSDEQLPLL
     101  EQLSGSGKAL  LVDRNGLYLA  NANFHHAAE  ELGLLAAEVA  QMEKKYRLXI
     151  KNNLYINNNA  WGVCDPSGQS  ELTFFPLYIG  STKFILVIGG  IPDLGKEAFV
     201  TLVRXLYXXL  QQPRVKLGRE  XGLCSNY*

```

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

```

      orfl43a.pep  MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSLKADADEMVSSEKLLTWAXTA
      orfl43-1    MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLHSLKADADEMVSSEKLLTWADTA

35      orfl43a.pep  DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
      orfl43-1    DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA

40      orfl43a.pep  NANFHHAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFFPLYIG
      orfl43-1    NANFHHAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPSGQSELTFFFPLYIG

      orfl43a.pep  STKFILVIGGIPDLGKEAFVTLVRXLY
45      orfl43-1    STKFILVIGGIPDLGKEAFVTLVRILY

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

```

50      orfl43.pep  MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLXDEQLPLLMEQL 60
      orfl43ng     MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQL 60

      orfl43.pep  SGSGKALLVDRNGLYLANANFHHAAEELGLLAAEVAQMEKKYRLLIKNN 110
55      orfl43ng     SGSGKALLVDRNGLYLANANFHHEAEELGLLAAEVAQMEKKYRLIRNNLYINNNAWGV 120

```

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

-347-

1 MRTKWSAVRS CSRADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLSD
 51 EQLPLLMEQL SGSGKALLVD RNgLYLANAN FHESAEELG LLAAEVAQME
 101 KKYRLLIRNN LYINNNAWGV CDPGQSELT FFPLYIGSTK FILVIAGIPD
 151 LSKGGICYFG KDFIPPLQOP RVKLGTGGIM RQLLISILED LNNTSTDIIA
 201 SAVISTDGLP MATMLPSHLN SDRVGAISAT LLALGSRVQ ELACGELEQV
 251 MIKKGSGYIL LSQAGKDAVL VLVAKETGRL GLILLDAKRA ARHIAEAI*

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

1 ATGGAATCAA CACTTTTCACT ACAAGCGAAT TTATATCCCT GCCTGACTCC
 51 TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCGGTAAAA
 101 CTTTGTTCG CAGCCTGTTG AAAGCGGATG CGGACGAAGT GGTGACAGT
 151 GAGAAGCTGC TCGCGGCGGA CACCGCCGAC ATCGATACCG CTTTGAACCT
 201 GTTGTACCGT TTGCAAAAAC TCGAATTCCT CTATGGCGAT GAAAACGGTC
 251 ATTCAGACGG CATCAATTTG TCGGACGAGC AATTGCCGTT GCTGATGGAA
 301 CAATTGTCCG GCAGCGGTAA GGCATTATTG GTCGATCGGA ACGGCTCTGT
 151 TCTTGCCAAC GCCAATTTCC ATCATGAGTC GCGGGAAGAG TTGGGGTTGT
 401 TGGCGGCAGA AGTCGCACAG ATGGAAAAGA AATACCGGCT GCTGATTAGG
 451 AACAACTGT ATATCAACAA TAACGCTTGG GCGTTTTCG ATCCTTCCGG
 501 TCAGAGCGAA TTGACATTTT TCCCATTGTA TATCGTTCA ACCAAATTTA
 551 TTTGGTTAT CGCGGCATT CCGATTGTA GCAAAGAGGC ATTTGTTACT
 201 TTGGTAAGGA TTTTATACCG CCGTTACAGC AACCGCGTGT AA

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

1 MESTLSLQAN LYPCLTPAGA FYAVSSDAPS AGKTLRLSLL KADADEVVSS
 51 EKLLAADTAD IDTALNLLYR LQKLEFLYGD ENGHSDGINL SDEQLPLLME
 101 QLSGSGKALL VDRNgLYLAN ANFHESAEEL LGLAAEVAQ MEKKYRLLIR
 251 NNLYINNNAW GVCDPSGQSE LTFFPLYIGS TKFILVIAGI PDLSKEAFVT
 201 LVRILYRRYS NRV*

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

orf143ng-1.pep MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSEKLLA-ADTA 59
 30 orf143-1 MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSEKLLTWADTA 60
 orf143ng-1.pep DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNgLYLA 119
 35 orf143-1 DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNgLYLA 120
 orf143ng-1.pep NANFHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFPLYIG 179
 orf143-1 NANFHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFPLYIG 180
 40 orf143ng-1.pep STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV 213
 orf143-1 STKFILVIGGIPDLGKEAFVTLVRILYRRYSNRV 214

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 50 51 GTTTCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGr
 101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
 151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTTC ACCGTGGTC
 201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG
 251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
 301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
 55 351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAA wTyCCAGCGT
 401 CCGTGGATG..

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLLQRLQGL ADNKICAFW FVRRFDEER VPQXAASMTF TLLALVPVL
51 TVMVAVASIF PVFDRWSDSF VSFVNQTIYP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM...
```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51 GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCCGAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTTC ACCGTGGTTC
201 GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGCGCGCG
251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
15 451 CTGTCTTTGG CGTGGGCAAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTCATGAC CTTTTGCTGT GGGGGCTGTA CCGCTTCGTG
601 CCAAACCGCT TCGTTCCCGC GCGGCAGGCG TTTGTGCGGG CTTTGCCAAC
651 AGCGTTTGTG TCGTAAACCG CGCGCTCCCT CTTCACTTGG TATATGGGCA
20 701 ATTTGACGCG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTTCAAG AGTTCAGACG
25 951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCT GTGGAAGGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30 1201 CAGGCGAAAA AACGGCAGTA G
```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLLQRLQGL ADNKICAFW FVRRFDEER VPQXAASMTF TLLALVPVL
51 TVMVAVASIF PVFDRWSDSF VSFVNQTIYP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
35 151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFT LLLWGLYRFV
201 PNRFPARQA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLAHGYIY SGRQGVWLKT
351 GADSIENEL FKLIFYRPLP VERDHVNQAV DAVMTPLQT LNMTLAEFDA
40 401 QAKKRQ*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N. meningitidis*:

```

45      10      20      30      40      50      60
orfl44.pep MTFLLQRLQGLADNKICAFWFVRRFDEERVVPQXAASMTFTTLLALVPVLTVMVAVASIF
orfl44a    MTFLLQRLQGLADNKICAFWFVRRFDEERVVPQXAASMTFTTLLALVPVLTVMVAVASIF
50      10      20      30      40      50      60
orfl44.pep PVFDRWSDSFVSFVNQTIYPXGADMVFDYINAFREQANRLTAIGSVMLVVTSMLLIRTID
orfl44a    PVFDRWSDSFVSFVNQTIYPQGADMVFDYINAFREQANRLTAIGSVMLVVTSXMLIRTID
55      70      80      90      100     110     120
orfl44.pep NTFNRIWRVXXQRPWM
orfl44a    NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
60
```

130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1   ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
5   51  GTTTGTCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTCC ACCGNTGGTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
251 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCG
10 351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAN CTTTCATGAC CTTTGTCTGT GGGGGCTGTA CCGCTNCGTG
15 601 CCAAACCGCT TCGTCCCGC GCGGCANGCG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTCGTG CTGGAACCG CGCGTTCCTT CTTTACTTGG TATATGGGCA
701 ATTTTCGACG CTACCGCTCG ATTTACGGNG CGTTTGCCCG CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
20 801 CGTGAATTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGNCT
851 TCGACTCGCG CGGACGTTT GACGACGTGT TGAATATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CNAAGCCTTG CCTGTTCAGG AGTTCAGACG
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACACTAT TCCGGCAGAC AGGGTTGGGT GTTGAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAATC TCAAGCTCT TCGTTTACCG
25 1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGATGCCGTG TTTGCAGCT TTGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1   MTFLLRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
30 51  TVMVAVASIF PVFDRWSDSF VSFVNQTIIV QGADMVFDYI NAFREQANRL
101 TAIGSVMLV TSXMLIRTID NTFNRIWRVN SQRPMWQFL VYWALLTFGP
151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
201 PNRFPARXA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFVAVPF
35 251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRFDSRGRF DDVLKILLLL
301 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
351 GADSIENEL FKL FVYRPLP VERDHNQAV DAVMMPCLQT LNMTLAEFDA
401 QAKKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

40 orf144a.pep MTFLLRLQGLADNKICAFW FVRRFDEERVPQAAASMTFTLLALVPVLTVMVAVASIF
    orf144-1 MTFLLRLQGLADNKICAFW FVRRFDEERVPQAAASMTFTLLALVPVLTVMVAVASIF

    orf144a.pep PVFDRWSDSFVSFVNQTIIVQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
45    orf144-1 PVFDRWSDSFVSFVNQTIIVQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID

    orf144a.pep NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
50    orf144-1 NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

    orf144a.pep RTAATLXFMTLLLWGLYRXV PNRFPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS
    orf144-1 RTAATLXFMTLLLWGLYRXV PNRFPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS

55    orf144a.pep IYGAFVAVPFFLLWLNLLWTLVLGGAVLTSLSYWQGEAFRRFDSRGRFDDVLKILLLL
    orf144-1 IYGAFVAVPFFLLWLNLLWTLVLGGAVLTSLSYWQGEAFRRFDSRGRFDDVLKILLLL

    orf144a.pep DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIENEL
60    orf144-1 DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIENEL

    orf144a.pep FKL FVYRPLPVERDHNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS 408
65    orf144-1 FKL FVYRPLPVERDHNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS 406

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKAFAWFVRRFDEERVQPXAAASMTFTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKAFAWFVIRRFSEERVQAAASMTFTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
15	orf144.pep	NTFNRIWRVXXQRPWM	136
	orf144ng	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

	1	MTFLQCWQGS	ADNKAFAW	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDQANRL
20	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPAQQA	FVGALITAF	LETARFLFTW	YMGNF	FDGYRS
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
25	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMT	PCLQT
	401	QAKKQQQS*				LNMTLAEFDA

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTTGCATGG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
30	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
	151	ACCGTAATGG	TCGCGGTCGC	TTCGATTTTC	CCCCTGTTCG	ACCGCTGGTC
	201	GGATTCTGTC	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CAGGGCGCGG
	251	ATATGGTGTG	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
35	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTCTG
	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCCTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCTC
	451	TTGTCTTTGG	GTGTGGGCAT	TTCTTTTATG	GTCGGGTCGG	TTCAAGACTC
	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACGCGTTG	AAGACGGCGG
40	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCGTG
	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGGCG	TTTGTCGGAG	CTTTGATTAC
	651	GGCATTCTGC	CTGGAGACGG	CACGTTTCCT	GTTACCTGG	TATATGGGCA
	701	ATTTCGACGG	CTACCGCTCG	ATTACGGCG	CATTGCGCG	CGTGCCGTTT
	751	TTCTGCTGT	GGTTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
45	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
	901	GATGCGGCGC	AAAAAGAAGG	CCGAACCTTG	TCCGTTTCAAG	AGTTCAGACG
	951	GCATATCAAT	ATGGGTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
50	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG
	1101	CCCGTTGCct	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAcgcctg	TTTGACAGCT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGgcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

55	1	MTFLQWQGL	ADNKAFAW	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPAQQA	FVGALITAF	LETARFLFTW	YMGNF	FDGYRS
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT

-351-

351 GADSIELSEL FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
401 QAKKQQQS*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5  orf144ng-1.pep MTFLQRWQGLADNKICAFWVIRRFSEERVPAQAASMTFTTLLALVPVLTVMVAVASIF
   orf144-1      MTFLQRWQGLADNKICAFWVIRRFSEERVPAQAASMTFTTLLALVPVLTVMVAVASIF
10 orf144ng-1.pep PVFDRWSDSFVSFVNQITIVPQADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
   orf144-1      PVFDRWSDSFVSFVNQITIVPQADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
15 orf144ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
   orf144-1      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
20 orf144ng-1.pep KTAARLAFMTLLLWGLYRFVFNRFVPAQAFVFGALITAFCLTARFLFTWYMGNFDGYRS
   orf144-1      RTAATLTFMTLLLWGLYRFVFNRFVPAQAFVFGALITAFCLTARSLFTWYMGNFDGYRS
25 orf144ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
   orf144-1      IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
30 orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSEL
   orf144-1      DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYYSGRQGWVLKTGADSIELNEL
   orf144ng-1.pep FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA QAKKQQQS
   orf144-1      FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA QAKKRQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1  ..AGACACGCC GCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
51 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
101 GCACCGATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
40 151 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
201 CCTGCTTGA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51 TRRWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

1  ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
51 CGAACGCTAC CGTACCGCC GCCTCATCCA CGCGTCCGG CTCGGCGGGG
101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
50 151 GAGTGGATAG GGATGACCGT CTTGTCGTC CTCGGCATGC TCCAGTTTCA
201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
401 CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
55 451 CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCCG

```

-352-

5 501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGCATGA CCCGCGAACG CCTCGAGGAG AACATGGCGA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGCAG CCATCTCGCC GCCACATCGG
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 10 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15 1 MNTSQRNRLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
 51 EWIGMTVFVU LGMLQFQGA IYSKAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCMLIG DNGSEWLD SG
 151 LMRAMNVLI GAAIAIAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
 201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
 20 251 RKIVNTTELL LTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRKWLDAHE RQHLRQSLLE TREHG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

	orf146.pep			10	20	30
				RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF				
		280 290 300 310 320 330				
	orf146.pep		40	50	60	70
			LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHLRQSLLETREHG			
35	orf146a	LWLSTNMRQEISALVILLQRTTRRKWLDAHERQHLRQSLLETREHSX				
		340 350 360 370				

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
 51 CGAACGCTAC CGTACCGCC GCCTCATCCA CGCCGTCCGG CTCGGCGGGG
 101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
 151 GAGTGGATAG GGATGACCGT CTTGCTCGTC CTCGGCATGC TCCAGTTTCA
 201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
 45 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
 301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA
 401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGTT CGACAGCGGC
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
 50 501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGCATGA CCCGCGAACG CCTCGAAGAG AACATGGCGA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGCAG CCACCTCGCC GCCACATCGG
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
 55 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
 60 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCGCTTGAA ACACGGGAAC ACGGCTGA

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

      1 MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
     51 EWIGMTVFV V LGMLQFOGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
    101 GNLLFYLTVG TASALAGWAA VGKNGYVPML AGLTMCMLIG DNGSEWFDSD
    151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG
    201 RRMTRERLEE NMAKMRQINA RMVKSRSULA ATSGESRISP AMMEAMQHAH
    251 RKIVNTTELL LTAAKLQSP KLNQSEIRLL DRHFTLLQTD LQQTVALING
    301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
    351 TRRKWLDAHE RQHLRQSLLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```

      orf146a.pep  MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
      orf146-1     MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
    15  orf146a.pep  LGMLQFOGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
      orf146-1     LGMLQFOGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
    20  orf146a.pep  VGKNGYVPMLAGLTMCLIGDNGSEWFDSDGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
      orf146-1     VGKNGYVPMLAGLTMCLIGDNGSEWFDSDGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
    25  orf146a.pep  FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSULAATSGESRISP
      orf146-1     FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSULAATSGESRISP
    30  orf146a.pep  AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNQSEIRLLDRHFTLLQTDLQQTVALING
      orf146-1     AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNQSEIRLLDRHFTLLQTDLQQTVALING
    35  orf146a.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE
      orf146-1     RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE
    40  orf146a.pep  RQHLRQSLLETREHSX
      orf146-1     RQHLRQSLLETREHG

```

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

      orf146.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF 30
      orf146ng   KLNQSEIRLLDRHFTLLQTDLQQTAAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364
    45  orf146.pep  LWLSTDMRQEISALVILLQRTRRKWLDAHERQHLRQSLLETREHG 75
      orf146ng   LWLSTNMRQEISALVIPLQTRRKWLDAHERQHLRQSLLETREHG 409

```

50 An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

      1 MSGVRFPSA PIPSTDPPSG SLCFFTFPLQ TASDMNSSQR KRLSGRWLNS
     51 YERYRHRRLI HAVRLGGTVL FATALARLLH LQHGEWIGMT VEVVLGMLQF
    101 QGAIYSNAVE RMLGTVIGL AGLGVLWLNQ HYFHGNLLFY LTIGTASALA
    151 GWAAGVGKNGY VPMLAGLTM MLIQDNGSEW LDSGLMRAMN VLIGAAIAIA
    201 AAKLLPLKST LMWRFMLADN LADCSKMAIE ISNGRRMTRE RLEQNMMVKMR
    251 QINARMVKSR SHLAATSGES RISPSMMEAM QHAHRKIVNT TELLTTAAK
    301 LQSPKLNQSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
    351 EALAEHLHYQ WQGF LWLSTN MRQEISALVI PLQTRRKWL DAHERQHLRQ
    401 SLLETREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

-354-

1 ATGAACTCCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
 51 CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCgga
 101 ccgtCCTGTT CGCCACCGCA CTCGCCCGgc tACTCCACCT CCAacacggc
 151 gAATGGATAG GGAtgaCCGT CTTCGTCGTC CTCGGCATGC TCCAGTTCCA
 5 201 AGGCgcgatt tActccaacg cggtgGAacg taTGctcggt acggtcatcg
 251 ggctgGGCGC GGGTTTGGgc gTTTATATGGC TGAACAGCA TTAtttccac
 301 ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccgg
 351 ctGGGCGGCG GTCCGCCAAA acggctacgt ccctatgctg GCGGGGctgA
 10 401 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC
 501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGTATGA CGCGCGAAGC TTTGGAGCAG AATATGGTCA AAATGCGCCA
 15 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG
 701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCCCAC
 751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
 801 CGAATCTCCC AAACCTCAACG GCAGCGAAAT CCGGCTGCTC GACCGCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAAGTGA
 20 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT AINPELEALA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACGCTGA

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

25 1 MNSSQRKRLS GRWLNSYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG
 51 EWIGMTVFVV LGMLQFOGAI YSNAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLD SG
 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
 201 RRMTRERLEQ NMVKMRQINA RMVKSRSHLA ATSGESRISP SMMEAMQHAH
 30 251 RKIVNTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTAAALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRWLDAHE RQHLRQSLLE TREHG*

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

35 orf146-1.pep MNTSQRNRLVSRWLNSYERYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
 orf146ng-1 MNSSQRKRLSGRWLNSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV
 orf146-1.pep LGMLQFOGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIVGTASALAGWAA
 40 orf146ng-1 LGMLQFOGAIYSNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
 orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
 45 orf146-1.pep FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
 orf146ng-1 FMLADNLADCSKMIAEISNGRRMTRERLEQNVMKMRQINARMVKSRSHLAATSGESRISP
 50 orf146-1.pep AMMEAMQHAHRKIVNTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
 orf146ng-1 SMMEAMQHAHRKIVNTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTAAALING
 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRWLDAHE
 55 orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRWLDAHE
 orf146-1.pep RQHLRQSLLETREHG
 60 orf146ng-1 RQHLRQSLLETREHG

Furthermore, ORF146ng-1 shows homology with a hypothetical *E. coli* protein:

65 sp|P33011|YEEA ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION
 >gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:o348#20; similar to [SwissProt
 Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)
 ORF_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

-355-

>gi1788318 (AE000292) f352; 100% identical to fragment YEEA_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352
Score = 109 bits (271), Expect = 2e-23
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

5 Query: 20 YRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAISNAVERML 79
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
Sbjct: 15 YRHYRIVHGTRVALAFLTLFLIIRLFTIPESTWPLVTMVIMGPISFWGNVVPRAFERIG 74

10 Query: 80 GTVIGLGAGLGVLWLNQHYFHNLLFYLTIGTASALAGWAAVGNKNGYVPMLAGLTMCMCLI 139
GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++
Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

15 Query: 140 GDNGSEWLDSGLMRAMNVLIGXXXXXXXXXXKLLPLKSTLMWRFMLADNLADCSKMI AEISN 199
G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +
Sbjct: 132 GSPTGE-IDTALWRSQDVLGSLLAMLEFTGIWPQRAFIHWRIQLAKSLTEYNRVYQSAFS 190

20 Query: 200 GRRMTRERLEQNVMVKMRQINARMVKSRSRLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259
+ R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V
Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRI PKSIYEGIQITNRNLVCMLEL 247

25 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXXAALINGRHARRIRIDTAINPEL 316
+ LN ++R D AL G +N +
Sbjct: 248 QINAYWATRP SHFVLLNAQKLR--DTQHMMQIILLSLVHALYEGNPQPVFANTEKLNDV 305

Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMQRQEISALVILLQRTTRK 354
E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLLNHHDLKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

35 1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA
51 GGGCAAATC GTCAGTGTGC GCGAACACAA CGAACGCGAG ATGGCGGACA
101 AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTCGGAT
151 CGGGGTACGC CGGCGGTGTG CGACCCGGGC GCGAACTCG CCCGCGCGT
201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA
40 251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATT TTATTCAAC
301 GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAACTGT TTGCCAAATG
351 GGTGCGGGCG GCGTTTCCTA TCGTCATGTT TGAAACGCGC CACCGCATCG
401 GTGCAGCGCT TGCCGATATG GCGGAACGT TCCCCGAACG CCGATTAATG
451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT
501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG
45 551 AGATGGTGT GGTGCTTTAT CCGCGCAGG ATGAAAAACA CGAAGGCTTG
601 TCCGAGTCCG CGCAAAACAT CATGAAATC CTCACAGCCG AGCTGCCGAC
651 CAAACAGGCG GCGGAGCTTG CTGCCAAAAT CACGGGCGAG GGAAGAAAG
701 CTTTGTACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

50 1 ..AEDTRVTAQL LSAYGIQKQL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51 AGTPAVCDPG AKLARRVREA GFKVVPVVA XAVMAALSA VEGSDFYFN
101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGAAALAD AELFPERRLM
151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL
201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKKALYD..

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTGGCGGAC ATTACCTGC
101 CGGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTGAGG GCAAACCTCG

-356-

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT
251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351 GTTTAAAGTC GTTCCCCTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTTCACGG TTTGTACCG
451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGGCGGC
501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCAGCGGAA
601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG
701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAAACATCA TGAAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801 GGAAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
851 TGGCTCTGTC TTGGAAAAAC AAATAG

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQGLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVA GVEG SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLM LARE
201 ITKTFTFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

25 Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
Orf286: 43 AEDTRHTGLLLQHFGINARL FALHDHNEQQKAETLLAKLQEGQNI ALVSDAGT PLINDPG 102

30 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLF AKWVRA 120
L R RE F + GF+P KS RR
Orf286: 103 YHLVRTCREAGIRVVPLPGPCA AITLSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAE 162

35 Orf147: 121 AFPIVMFETPHRIG AALADMAELFPERR-LMLAREITKTFTFLSGTVGEIQTALSADGD 179
++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
Orf286: 163 PRTLIFYESTHRLLDSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDEN 222

40 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALY 236
+ +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAE LPLKAAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45 orf147.pep AEDTRVTAQLLSAYGIQGLVSVREHNERQ
|||
orf75a TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGLVSVREHNERQ
20 30 40 50 60 70

50 orf147.pep MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA
|||
orf75a MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA
80 90 100 110 120 130

55 orf147.pep GVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFPIVMFETPHRIG AALADMAELFPERRLM
|||
orf75a GVAGSDFYFNGFVPPKSGERRKLF AKWVRAAFPIVMFETPHRIG ATLADMAELFPERRLM
140 150 160 170 180 190

60 orf147.pep LAREITKTFTFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI
160 170 180 190 200 210

-357-

```

|||||
orf75a  LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI
      200      210      220      230      240      250

      220      230
orf147.pep LTAE LPTKQAAELAAKITGEGKKALYD
|||||
orf75a  LTAE LPTKQAAELAAKITGEGKKALYDLALS WKNKX
      260      270      280      290

```

10 ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from *N. gonorrhoeae*:

Line	Sequence	Position
15	orf147.pep	AEDTRVTAQLLSAYGIQGKLVSVREHNERQ 30
	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGRVSVREHNERQ 85
	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA 90
20	orf147ng	MADKVIGFSLDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA 145
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFKQWVRAAFPIVMFETPHRIGAAALADMAELFPERRLM 150
25	orf147ng	GVAESDFYFNGFVPPKSGERRKLFKQWVRAAFPVVMFETPHRIGATLADMAELFPERRLM 205
	orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLVYPAQDEKHEGLSESAQNIMKI 210
	orf147ng	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLVYPAQDEKHEGLSESAQNAMKI 265
30	orf147.pep	LTAEPLTKQAAELAAKITGEGKKALYD 237
	orf147ng	LAAELPTKQAAELAAKITGEGKKALYDLALSWNK 300

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

```

35      1  MSVFQTAFFM  FQKHLQKASD  SVVGGTLYVV  ATPIGNLADI  TLRALAVLQK
      51  ADIICAEDTR  VTAQLLSAYG  IQGRLVSVRE  HNERQMAADV  IGFLSDGLVV
     101  AQVSDAGTPT  VCDPGAKLAR  VRREAGFKVV  PVVGASADVM  ALSVAGVAES
     151  DFYFNGFVPP  KSGERRKLFA  KQVRAAFPVV  MFETPHRIGA  TLADMAELFP
     201  ERRMLLAREI  TKTFETFLSG  TVGEIQTALA  ADGNQSRGEM  VLVLYPAQDE
40     251  KHEGLSESAQ  NAMKILAAEL  PTKQAAELAA  KITGEGKKAL  YDLALSWKNK
     301  *

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
45	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
	101	CGCGTTTGGC	GGTATTGCAA	AAGGCGGCAG	TCAATTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GCGATTCAAG	GCAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
50	251	TCCTTTTACA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
55	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGCG
	501	ATTTCTGTCT	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGGCTTG
	551	CCGATATGGC	GGAATTGTTC	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
	601	ATCACGAAAA	CGTTTGAAC	GTTCTTAAAG	GGCACGGTTG	GGGAAATTCA
60	651	GACGGCATTG	GCGGCGGACG	GCAACCAAT	GCGCGCGCAG	ATGTTGTTGG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751	CAAAATGCGA	TGAAAATCCT	TGCGGCGCAG	CTGCCGACCA	AGCAGGCGGC
	801	GGAGCTTGCC	GCCAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	TGGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
      51 RVTAQLLSAY GIQGRVLSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
    101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
      151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLDMAELF PERRLMLARE
    201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
      251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

ORF147ng shows homology to a hypothetical *E.coli* protein:

```

    sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    >gi|606086 (U18997) ORF_f286 [Escherichia coli]
    >gi|1789535 (AE000395) Hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
    Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

    Query: 4   KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
              K  Q  A +S   G  LY+V TPIGNLADIT RAL VLQ  D+I AEDTR T  LL  +GI
    Sbjct: 2   KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLAQVDLIAAEDTRHTGLLLQHFGIN 59

    Query: 64   GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
              RL ++ +HNE+Q A+ ++  L +G  +A VSDAGTP + DPG  L R  REAG +VVP+
    Sbjct: 60   ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

    Query: 124  VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFPVVMFETPHRIGATL 183
              G  A  + ALS AG+   F  + GF+P KS  RR              ++ +E+ HR+  +L
    Sbjct: 120  PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAEPRTLIFYESTHRLDLSL 179

    Query: 184  ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
              D+  +  E R ++LARE+TKT+ET   VGE+  +  D N+ +GEMVL++  +
    Sbjct: 180  EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

    Query: 243  HEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLAL 286
              E  L  A  + +L AELP K+AA LAA+I G  K ALY  AL
    Sbjct: 239  EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

      1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51 AACCGGTCGC ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
    101 TCGGCATTCT TCCCAAGGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
      151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
    45 201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
      251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
      301 GTGGCGGcAT TGGTGGGCGt ATCAATATAT TGTGAGCGTG GCACATAACG
      351 GCGGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAak.AA tATCCC.GAT
      401 CAACAwCGwW TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
    50 451 GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGCATA
      501 AATWTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG
      551 CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTT GTATTGGGGC
      601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACC GCGAAA
      651 GTTCATATCA TATTGCAAGT .....
    701 ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
      751 AAAGTGTTA ATTAATGGGG TATTGCAAAAC GGGCAACCCC TATATAGGAA
      801 AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
      851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACACGTC AAAATGGGAA
      901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC

```

-359-

951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC
 1051 AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAAATATTT
 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT
 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCTGA
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA
 1251 CCGTTACTTG GAAAGTAAAC GGCCTGGCAA ACGACCGCCT GTCCAAATC
 1301 GGCAAAGGCA CGCTG.....
 2101 GATAAAG
 2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC
 2201 GATCACGCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT
 2251 TAGTGCAAAT GGCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA
 2301 ACGGCAACCK TA_gCcTCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC
 2351 ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCGGC AACGCTAAGG
 2451 CAAACGTAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG
 2501 GCAGTATTCC ATTTTGAAG CAGCCGCTT ACCGACAAA TCAGCGCGCG
 2551 CAAGGATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCCGTCa
 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TAcACTAAGT
 2651 TCCGCCATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA
 2701 TGCGCGCGCG CGCGTTCGCG GCCGTTCGCG CCGTTCCTTA TTATmCGTTA
 2751 CACCGCCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAAACGGC
 2801 AAATTGAACG GTCAGGGAAC ATTCCGCTT ATGTCGGAAC TCTTCGCTA
 2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT
 2901 TGGCGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG
 2951 GTAGTGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC
 3001 CCTGCAAAAC GAACACGTCG ATGCAGGCGG GTGG.....
 3551 TTAGAC CGCGTATTTG CCGAAGACCG
 3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAA CACTACCGTT
 3651 CGCAAGATT CCGCGCTAC CGCCAACAAA CCGACCTGCG CCAAATCGGT
 3701 ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGCATCCTGT TTTGCGACAA
 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG
 3801 CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC
 3851 ATCAG_gCGCG GCGCGGGTT TTAGCAGCGG CAGCCTTtCa GACGGCATCG
 3901 GAGsmAAAwT CCGCGCGCGC GTGctGCATT ACGGCATTCa GGCACGAtAC
 3951 CGCGCCGgtt tCg_gCGgAtt CCGCATCGAA CCGCACATCG GCGCAACCGC
 4001 ctATTTGCTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA
 4051 CCCCCGCGCT TGCATTCAAC CGcTACCGCG CGGGCATTAA GGCAGATTAT
 4101 TCATTCAAAC CGGCGCAACA CATTTCCATC ACGCCTTATT TGAGCCTGTC
 4151 CTATACCGAT GCCGCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCGC
 4201 TATTGGCTCA GGATTTTCGG AAAACCGCA GTGCGGAATG GGgCGTAAAC
 4251 GCCGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCAAAGG
 4301 CCGCAACTG GAAGCGCAAC ACAGCGCGG CATCAAATTA GGCTACCGCT
 4351 GGTA...
 701 DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS
 801 DHAVQNGSLT LSGNAKANVS HSALNGNVSL ADKAVFHFES SRFTGQISGG
 851 KDTALHLKDS EWTLP_gSGXEL GNLNLDNATI TLNSAYRHTA AGAQTGSATD
 901 APRRRSR_gRSR RSL_gLXVTPPT SVESRFNTLT VNGKLNGQGT FRFMSELFY
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLFNT
 1001 LQNEHVDAGA W.....
 1151 LDRVFAEDR
 1201 RNAVWTS_gGIR DTKHYRSQDF RAYRQOTDLR QIGMQKNLGS GRVGILFSHN
 1251 RTENTFDDGI GNSARLAHGA VFGQY_gIDRF YIGISAGAGF SSGSLSDGIG
 1301 XKXRRRV_gLHY GIQARYRAGF GGFGIEPHIG ATRYFVQKAD YRYENVNIAT
 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDASG KVRTRVNTAV

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

1 MKTTDKRTE THRKAPKTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN
 51 YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS_gRNG
 101 VAALVG_gVQYI VSVAHNGGYN NVDFGAEGXN IXDQXR_gTYK IVKRNNYKAG
 151 TKGHPYGGDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYPDRVRIGA
 201 GRQYWRSD_gED EPNNRESSYH IAS.....GS PMFIYDAQKQ
 251 KWLINGVLQT GNPYIGKSNG FQLVRKDWFY DEIFAGDTHS VFYEPRONGK
 301 YSFNDNNGT GKINAKHEHN SLPNRLKTRT VQLFNVSLSE TAREPVYHAA
 351 GGVNSYRPR_gL NNGENISFID EGKGELILTS NINQGAGGLY FQGDFTVSPE
 401 NNETWQAGV HISEDSTVTW KVN_gVANDRL SKIGKGT_gL.....
 701 DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS
 801 DHAVQNGSLT LSGNAKANVS HSALNGNVSL ADKAVFHFES SRFTGQISGG
 851 KDTALHLKDS EWTLP_gSGXEL GNLNLDNATI TLNSAYRHTA AGAQTGSATD
 901 APRRRSR_gRSR RSL_gLXVTPPT SVESRFNTLT VNGKLNGQGT FRFMSELFY
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLFNT
 1001 LQNEHVDAGA W.....
 1151 LDRVFAEDR
 1201 RNAVWTS_gGIR DTKHYRSQDF RAYRQOTDLR QIGMQKNLGS GRVGILFSHN
 1251 RTENTFDDGI GNSARLAHGA VFGQY_gIDRF YIGISAGAGF SSGSLSDGIG
 1301 XKXRRRV_gLHY GIQARYRAGF GGFGIEPHIG ATRYFVQKAD YRYENVNIAT
 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDASG KVRTRVNTAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW
1451 *

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

5 1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCGT
101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
201 GCGGAAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAAT
251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTG CCGTAACGGC
10 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
401 ATCGTTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT
451 AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCATAAATT
501 TGTACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
15 551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCGTAT TGGGCGAGGC
601 AGGCAATATT GGCGATCTGA TGAAGATGAG CCCAATAACC GCGAAAAGTTC
651 ATATCATATT GCAAGTGCGT ATTCTTGGCT CGTTGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGAAAAAATT
20 751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
801 TGGCTCACCA ATGTTTATCT ATGATGCCCC AAAGCAAAAG TGGTTAATTA
851 ATGGGGTATT GCAACGGGGC AACCCTTATA TAGGAAAAAG CAATGGCTTC
901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTCAGTA TTCTACGAAC CAGTCAAAA TGGGAAATAC TCTTTTAAACG
1001 ACGATAATAA TGGCAGAGGA AAAATCAATG CCAACATGA ACACAATTCT
25 1051 CTGCCTAATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTTT TATTGACGAA
1201 GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
1251 ATTTATTTTC CAAGGAGATT TTACGGTCTC GCCTGAAAAT AACGAAACTT
30 1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGAAG ACAGTACCGT TACTTGAAAA
1351 GTAAACGGCG TGGCAACGGA CCGCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
1451 GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAACCAAGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGTACGGTGC AACTGAATGC
35 1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTTCGGCTTT CGCGGCGGAC
1601 GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTGCTAC AACCAGCAAT AACAACAGCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
40 1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAGC CCGCCGAGAG
1851 AGACCGCACC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA
1901 CGCAAAACAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
1951 TACAATCATT TAAACGACCA TTGGTCGCAA AAAGAGGGCA TTCCTCGCGG
2001 GGAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAAGCGGAAA
45 2051 ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCCGCAA TGTTGCCAAA
2101 GTGAAAGGCG ATTGGCATTT GAGCAATCAC GCCCAAGCAG TTTTGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAATTG TGTGCAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGACTAAGA CCGACATCAG CCGCAATGTC GATCTTGCCG ATCAGCGTCA
50 2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAATG
2351 GCGATACACG TTATACAGTC AGCCACAACG CCACCAAAA CGGCAACCTT
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
55 2501 TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC
2551 CATTCCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCAGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCTATCG
60 2751 CCACGATGCG GCAGGGGCGC AAACCGGCAG TCGACAGAT GCGCCGCGCC
2801 GCCGTTTCGG CCGTTCGCGC CGTTCCTTAT TATCCGTAC ACCGCCAAT
2851 TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
2901 TCAGGGAACA TTCCGCTTTA TGTGGAACCT CTTCCGCTAC CGCAGCACA
2951 AATTGAAGCT GCGGGAAGT TCCGAAGGCA CTTACACCTT GCGGTCAAC
3001 AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGGAAAG
65 3051 AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTACCC CTGCAAAACG
3101 AACACGTCGA ATCCGGCGCG TGGCGTTACC AACTCATCCG CAAAGACGGC
3151 GAGTTCCGCC TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAAACT
3201 CGGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAC GCGCAAGGCC
3251 TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAA GACAGAAAGC
70 3301 GTTGCCGAAC CGGCCGCGCA GGCAGGCGGG GAAAATGTG GCATTATGCA

-361-

5
10
15
20

```

3351 GCGCGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCCTTG
3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTTCCCCCGC
3451 GCGCGCCGCG CCCGCCGGGA TTTGCCGCAA CTGCAACCCC AACCGCAGCC
3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTTGAGTG
3551 AATTTTCCGC CACGCTCAAC AGCGTTTTTCG CCGTACAGGA CGAATTAGAC
3601 CGCGTATTTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651 GGACACCAAA CACTACCGTT CGCAAGATT CCGCGCCTAC CGCCAACAAA
3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC
3751 GGCATCCTGT TTTCGCACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
3801 CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTTCGG CAATACGGCA
3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGCG
3901 AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCGCCGCGG TGCTGCATTA
3951 CGGCATTAG GCACGATACC GCGCCGGTTT CGGCGGATTC GGCATCGAAC
4001 CGCACATCGG CGCAACGCGC TATTTTCGTCC AAAAAGCGGA TTACCGCTAC
4051 GAAAACGTCA ATATCGCCAC CCCCAGCCTT GCATTCAACC GCTACCGCGC
4101 GGGCATTAA GAGATTATT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4151 CGCCTTATTT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTTCGGG AAACCCGCGC
4251 TCGGGAATGG GCGGTAAACG CCGAAATCAA AGGTTTCACG CTGTCCCTCC
4301 ACCTGCGCGC CGCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351 ATCAAATTAG GCTACCGCTG GTAA

```

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

25
30
35
40
45
50

```

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDFAE NKGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS RNG
101 VAALVG DQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
151 KGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
251 KHSPYGF LPT GSGFDGSGSP MFIYDAQKQK WLVINGVLQTG NPYIGKSNFG
301 QLVKRDW FID EIFAGDTHSV FYEPRQNGKY SFNDNNNGTG KINAKHEHNS
351 LPNRLKTR TV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE
401 GKGE LILTSN INQGAGGLYF QGDFTVSPEN NETWQAGVH ISEDSTVTWK
451 VNGVANDRL S KIGKGT LHVQ AKGENQGSIS VGDGTVILDQ QADDKGGKQA
501 FSEIGLVSG R GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
551 DEGAMIVNH N QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGEKD
601 TTKTNGRLNL VYQPAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLNDHWS Q KEGIPRGEIV WDNDWINRTF KAENFOIKGG QAVVSRNVAK
701 VKGDWHL SNH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
751 LTKTDISGN V DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATNGNL
801 SLVGNAQAT F NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
851 HSA LNGVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSTGTEL
901 GNLNL DNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
951 SVESRENTLT VNGKLNGQGT FRMSELFY RSDKLKLAES SEGTYTLAVN
1001 NTGNEPASLE QLTVVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG
1051 EFRLHNPVKE QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101 VAEPARQAGG ENVGIMQAE EKKRVQADKD TALAKQREAE TRPATTAFFR
1151 ARRARRDL P LQPOPQPOPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
1201 RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQOTDLRQIG MQKNLGSGRV
1251 GILF SHNRTE NTFDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
1301 SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYR
1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAASGKVR
1401 TRVNTAVLA Q DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
1451 IKLGYRW*

```

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

60

```

          10      20      30      40      50      60
orfl.pep MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGIN YQYYRDFAE
          10      20      30      40      50      60
orfla    MKTTDKRTTETHRKAPKTGRIRFSPAYLAI CLSFGILPQAWAGHTYFGIN YQYYRDFAE
          10      20      30      40      50      60
          70      80      90     100     110     120
orfl.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVG VQYIVSAHNGGYN

```

-362-

BNSDOCID: <WO 9924578A2 1 >

-363-

```
|||||
orfla  GDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNGSLTSD
      780      790      800      810      820      830

5      550      560      570      580      590      600
orfl.ppep NAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGNL
      |||||
orfla  NAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSGTELGNL
      840      850      860      870      880      890

10     610      620      630      640      650      660
orfl.ppep NLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRLXVTPPTSVESRFNTLTVNG
      |||||
orfla  NLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVESRFNTLTVNG
      900      910      920      930      940      950

15     670      680      690      700      710      720
orfl.ppep KLNQGQTFRFMSELFGYRSDKLKLAESSEGTYYTLAVNNTGNEPASLEQLTVVEGKDNKPL
      |||||
orfla  KLNQGQTFRFMSELFGYRSDKLKLAESSEGTYYTLAVNNTGNEPVSLDQLTVVEGKDNKPL
      960      970      980      990      1000     1010

20     730      740      750
orfl.ppep SENLNFTLQNEHVDAGAW-----
25     |||||
orfla  SENLNFTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLGKAEAKKQAEKDNAQS
      1020     1030     1040     1050     1060     1070

30     -----
orfl.ppep -----
orfla  LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKRVQADKDSALAKQREAEATRP
      1080     1090     1100     1110     1120     1130

35     -----
orfl.ppep -----
orfla  XTTAFPRARXARRDLQPQPQPQPQDLXSRYANSGLSEFSATLNSVFAVQDELDR
      1140     1150     1160     1170     1180     1190

40     770      780      790      800      810      820
orfl.ppep VFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTEN
      |||||
orfla  VFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTEN
      1200     1210     1220     1230     1240     1250

45     830      840      850      860      870      880
orfl.ppep TFDDGIGNSARLAHGAVFGQYGRFYIGISAGAGFSSGSLSDGIGXKXRRRVLHYGIAQ
      :|||
orfla  XFDDGIGNSARLAHGAVFGQYGRFDIGISTGAGFSSGXLSDGIGGKIRRRVLHYGIAQ
      1260     1270     1280     1290     1300     1310

50     890      900      910      920      930      940
orfl.ppep RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
      |||||
orfla  RYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX
      1320     1330     1340     1350     1360     1370

55     950      960      970      980      990      1000
orfl.ppep SITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKP
      |||||
orfla  SITPYXLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHAAAAGKP
      1380     1390     1400     1410     1420     1430

60     1010     1020
orfl.ppep QLEAQHSAGIKLGYRWX
      |||||
orfla  QLEAQHSAGIKLGYRWX
      1440     1450

65

70     The complete length ORF1a nucleotide sequence <SEQ ID 651> is:
```

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCTG
	101	TCGGCATTCT	TCCCCAAGCT	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
5	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGTNT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AACGTTGATT	TGGTGGCGGA	AGGAAGNAAT	CCCCATCAGC
10	401	ACCGTTTTTC	TTACCAAAT	GTGAAAAGAA	ATAATTATTA	GCCTGACAAAT
	451	TCACACCTT	ACAACGGCGA	TTANCATATG	CCGCGTTTGC	ATAAATTTGT
	501	CACAGATGCA	GAACCTGTCT	AAATGACGAG	TGACATGAGG	GGGAATACCT
	551	ATTCCGATAA	AGAAAAATAT	CCCGAGCGTG	TCCGCATCGG	CTCAGGACAC
	601	CATATTGGC	GTTATGATGA	TGACAAACAC	GGCGATTAT	CCTACGCAAT
	651	CGCATGGTTA	ATTGGCGGCA	ATACACATAT	GCAGGGTTGG	GGAAATAATG
15	701	GCGTANTTAG	TTTGAGCGGC	GATGTGCGCC	ATGCCAACGA	CTATGCCCC
	751	ATGCCGATTG	CAGGTGCGGC	AGGCGACAGC	GGTTCGCCAA	TGTTTATTTA
	801	TGACAAAACA	AACAATAAAT	GGCTGCTCAA	CGGAGTTTAA	CAAACCGGCT
	851	ACCCTTATTC	CGGCAGGGAA	AACGGTTTCC	AGCTGATACG	CAAAGATTGG
	901	TTCTACGATG	ACATTTACAG	AGGCGATACA	CATACCGTCT	NTTTTGAACC
20	951	GCGCAGTAAC	GGACATTTT	CCTTTACATC	CAACAACAAC	GGTACGGGTA
	1001	CGGTAAACGA	AACCAACGAA	AAGGTNTCCA	ATCCAAAGCT	TAAAGTACAG
	1051	ACAGTCCGAC	TGTTTGACGA	ATCTTTGAAT	GAAACTGATA	AAGAACCAGT
	1101	TTACGCGGCA	GGGGGTGTTA	ATCAGTACCG	TCCAAGGTTA	AACAACGGTG
	1151	AAAACCTTTC	TTTTATCGAT	TACGGCAACG	GCAAACCTCAT	CTTATCAAAC
25	1201	AACATCAACC	AAGCGCGGG	CGGTTTGAT	TTTGAAGGTG	ATTTTACGGT
	1251	CTCGCTGAA	AACAACGAAA	CGTGGCAAGG	CGCGGGCGTT	CATATCAGTG
	1301	AAGACAGTAC	CGTTACTTGG	AAAGTAAACG	GCGTGGCAAA	CGACCGCCTG
	1351	TCCAAAATCG	GCAAAGGCAC	GCTGCACGTT	CAAGCCAAAG	GGGAAACCCA
	1401	AGGCTCGATC	AGCGTGGGCG	ACGGTACAGT	CATTTTGGAT	CAGCAGGCAG
30	1451	ACGATAAAGG	CAAAAAACAA	GCCTTTAGTG	AAATCGGCTT	GNTCAGCGGC
	1501	AGGGGTACGG	TGCAACTGAA	TGCCGATAAT	CAGTTCAACC	CCGACAAACT
	1551	CTATTTCCGG	TTTCGCGGCG	GACGTTTGGA	TTTAAACGGG	CATTCGCTTT
	1601	CGTTCCACCG	TATTCAAAAT	ACCGATGAAG	GGGCGATGAT	TGNCNATCAT
	1651	AATGCCACAA	CAACATCCAC	CGTTACCATT	ACAGGGAATG	AAAGTATTAC
35	1701	ACAACCGAGT	GGTAAGAATA	TCAATAGACT	TAATTACAGC	AAAGAAATTG
	1751	CCTACAACGG	TTGGTTTGGC	GAGAAAGATA	CGACCAAAAC	GAACGGCGGG
	1801	CTCAACCTTG	TTTACCAGCC	CGCCGCAGAA	GACCGCACCC	NGCTGCTTTC
	1851	CGGCGGAACA	AATTTAAACG	GCAACATCAC	GCAAACAAC	GGCAAACCTG
	1901	TTTTCAGCGG	CAGACCGACA	CCGCACGCC	ACAATCATTT	AGGAAGCGGG
40	1951	TGGTCAAAAA	TGGAAGGTAT	CCCACAAGGA	GAAATCGTGT	GGGACAACGA
	2001	CTGGATCNAC	CGCACGTTTA	AAGCGGAAAA	TTTCCATATT	CAGGGCGGGC
	2051	AGGCGGTGAT	TTCCCGCAAT	GTTGCCAAAG	TGGAAGGCGA	TTGNCATTTG
	2101	AGCAATCACG	CCCAAGCAGT	TTTGGTGTC	GCACCGCATC	AAAGCCATAC
	2151	AATCTGTACA	CTGTCGACT	GGACNGGTCT	GACAAATTGT	GTCGAANAAA
45	2201	NCATTACCGA	CGATAAAGTG	ATTGCTTCAT	TGACTAAGAC	NGACNTNAGC
	2251	GGCANTGTNA	GNCTNNCCNA	TNACGNTNNT	TNAAANCTCN	CNGGGCNTGC
	2301	NNCACTNAAN	GGCAATCTTA	GTGCAAATGG	CGATACACGT	TATACAGTCA
	2351	GCCACAACGC	CACCAAAAC	GGCAACCTTA	GCCTCGTGGG	CAATGCCCAA
	2401	GCAACATTTA	ATCAAGCCAC	ATTAACGGC	AACNCATCGG	NTTCGGGCAA
50	2451	TGCTTCATTT	AATCTAAGCA	ACAACGCCGC	ACAAAACGGC	AGTCTGACGC
	2501	TTTCCGACAA	CGCTAAGGCA	AACGTAAGCC	ATTCCGCACT	CAACGGCAAT
	2551	GTCTCCCTAG	CCGATAAGGC	AGTATTCCAT	TTTGAAAACA	CCCGCTTTAC
	2601	CGGACAACCT	AGCGGCAGCA	AGGANACAGC	ATTACACTTA	AAAGACAGCG
	2651	AATGGACGCT	GCCGTACGGC	ACGGAATTAG	GCAATTTAAA	CCTTGACAAC
55	2701	GCCACCATTA	CACTCAATTC	CGCCTATCGC	CACGATGCTG	CAGGCGCGCA
	2751	AACCGGCAGN	GTGTACAGCA	CGCCGCGCCG	CCGTTGCGCG	CGTTCCCTAT
	2801	TATCCGTTAC	ACCGCCAAT	TCCGTTAGAA	CCCGTTTCAA	CACGCTGACG
	2851	GTAAACGGCA	AATTGAACNG	TCAAGGAACA	TTCCGCTTTA	TGTCGGAAT
	2901	CTTCGGCTAC	CGAAGCGACA	AATTGAAGCT	GGCGGAAAGT	TCCGAAGGNA
60	2951	CTTACACCTT	GGCGGTCAAC	AATACCGGCA	ACGAACCCGT	AAGCCTCGAT
	3001	CAATTGACGG	TAGTGGAAGG	GAAAGACAAC	AAACCGCTGT	CCGAAAACCT
	3051	TAATTTTACC	CTGCAAAACG	AACACGTCGA	TGCCGGCGCG	TGGCGTTACC
	3101	AACTCATCCG	CAAAGACGGC	GAGTTCCGCC	TGCATAATCC	GGTCAAAGAA
	3151	CAAGAGCTTT	CCGACAAACT	CGGCAAGGCA	GAAGCCAAAA	AACAGGCGGA
65	3201	AAAAGACAAC	GCGCAAAGCC	TTGACGCGCT	GATTGCGGCC	GGGCGCGATG
	3251	CCGCCGAAAA	GACAGAAAGC	GTTGCCGAAC	CGGCCCGGCN	GGCAGGCGGG
	3301	GAAATGTCTG	GCATTATGCA	GGCGGAGGAA	GAGAAAAAAC	GGGTCAAGGC
	3351	GGATAAAGAC	AGCGCNTTGG	CGAAACAGCG	CGAAGCGGAA	ACCCGGCCGG
	3401	NTACCACCGC	CTTCCCCCGC	GCCCCGNGCG	CCCGCCGGGA	TTTGCCGCAA
	3451	CCGCAGCCCC	AACCGCAACC	TCAACCCCAA	CCGCAGCGCG	ACCTGATNAG
70	3501	CCGTTATGCC	AATAGCGGTT	TGAGTGAATT	TTCCGCGACG	CTCAACAGCG
	3551	TTTTCGCCGT	ACAGGACGAA	TTGGACCGCG	TGTTTGCCGA	AGACCGCCGC

-365-

5
10
15

```

3601 AACGCNGTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTCCGCA
3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC
3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTC GCACAACCGG
3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTTCGAC ATCGGCATCA
3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACGG CATCGGAGGC
3901 AAAATCCGCC GCCCGGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC
3951 CGGTTTCGGC GGATTTCGGCA TCGAACCGTA CATCGGCGCA ACGCGCTATT
4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTCAAT
4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCCTATA
4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG
4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGGC GAATGGGGCG TAAACGCCGA
4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGCC AAAGNCCGCG
4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 652>:

20
25
30
35
40
45

```

1  MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51  YQYYRDFAEF NGKFAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVSRNG
101 VAALVGDQYI VSVAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN
151 SHPYNGDXHM PRLHKFVTDA EPVEMTSDMR GNTYSDKEYK PERVERIGSGH
201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP
251 MPIAGAAGDS GSPMFIYDKT NNKWLNLGVL QTGYPSYGRE NGFLIRKDW
301 FYDDIYRGDT HTVXFEPASN GHFSFTSNNN GTGTVTETNE KVSNPKLKVQ
351 TVRLFDES LN ETDKEPVYAA GGVNQYRPRL NNGENLSFID YGNGKLILSN
401 NINQGAGGLY FEGDFTVSPE NNETWQAGAV HISEDSTVTW KVNGVANDRL
451 SKIGKGT LHV QAKGENQGS I SVGDGTVILD QQADDKGKKQ AFSEIGLXSG
501 RGTVQLNADN QFNPDKLYFG FRGGRLDLNG HSLSFHRIQN TDEGAMIXXH
551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
601 LNLVYQPAAE DRTXLLSGGT NLGNITQTN GKLFSSGRPT PHAYNHLGSG
651 WSKMEGIPOG EIVWDNDWIX RTFKAENFHI OGGQAVISRN VAKVEGDHXL
701 SNHAQAVFGV APHQSHITCT RSDWTGLTNC VEXXITDDKV IASLTKTDXS
751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATON GNLSLVGNAQ
801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNGN
851 VSLADKAVFH FENSRTGQL SSGKXTALHL KDEWTLPSG TELGNLNDN
901 ATITLNSAYR HDAAGAQTGX VSDTPRRRSR RSLLSVTPPT SVESRFNTLT
951 VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
1001 QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG EFRLHNPVKE
1051 QELSDKLKGA EAKKQAEKDN AQSLDALIAA GRDAAEKTES VAEPARXAGG
1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPXTTAFPR ARXARRDLPO
1151 PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
1201 NAVWTSXIRX TKHYRSQDFR AYRQOTDLRQ IGMQKNLGS RVGILFHSNR
1251 TENXFDDGIG NSARLAHGAV FGQYIGIRFD IGISTGAGFS SGXLSDGIGG
1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENVNIATP
1351 GLAENRYRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL
1401 AQDFGKTRSA EWGVNAEIKG FTLXSHAAAA KGPQLEAQS AGIKLGIRW*

```

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

50
55
60
65

```

          10      20      30      40      50      60
orfla.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAEF
orfl-1    MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAEF
          10      20      30      40      50      60

          70      80      90      100     110     120
orfla.pep KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN
orfl-1    KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN
          70      80      90      100     110     120

          130     140     150     160     170     179
orfla.pep NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTDAEPVEMTSDM
orfl-1    NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSY
          130     140     150     160     170     180

```

-366-

		180	190	200	210	220	230
	orfla.pep	RGNTYSDKEKYPERVIRIGSGHHYWRYYYYDKHGDLD--SYSGA---WLIGGNTHMQGWGNN					
	orfl-1	DGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG					
5		190	200	210	220	230	240
	orfla.pep	240	250	260	270	280	290
	orfl-1	GVXSLSGD--VRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLNGVLQGTGYPYSGRENG					
10		250	260	270	280	290	
	orfla.pep	GTVNLGSEKIKHS--PYGFLPTGGSGFGDSGSPMFIYDAQKQKWLINGVLQGTGNPYIGKSNG					
	orfl-1	GTVNLGSEKIKHS--PYGFLPTGGSGFGDSGSPMFIYDAQKQKWLINGVLQGTGNPYIGKSNG					
		300	310	320	330	340	350
	orfla.pep	FQLIRKDWFYDDIYRGDTHTVXFEPNSNGHFSFTSNNNGTGTVTETNEKVSNP--KLKVQT					
15		300	310	320	330	340	350
	orfl-1	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTKINAKHEHNSLPNRLKTRT					
		360	370	380	390	400	410
	orfla.pep	VRLFDESINETDKEPVY--AAGGVNQYRPRLNNGENLSFIDYNGKILSNNINQAGAGGLY					
20		360	370	380	390	400	410
	orfl-1	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY					
		420	430	440	450	460	470
	orfla.pep	FEGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTGLHVQAKGENQGS					
25		420	430	440	450	460	470
	orfl-1	FQGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTGLHVQAKGENQGS					
		480	490	500	510	520	530
	orfla.pep	SVGDGTVILDQQADDKGGKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG					
30		480	490	500	510	520	530
	orfl-1	SVGDGTVILDQQADDKGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG					
		540	550	560	570	580	590
	orfla.pep	HSLSFHRIQNTDEGAMIXXHNATTTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFG					
35		540	550	560	570	580	590
	orfl-1	HSLSFHRIQNTDEGAMIVNHQDKESTVTITGNKDIAT--TGNN--NSLDSKKEIAYNGWFG					
		600	610	620	630	640	650
	orfla.pep	EKDTTKTNGRNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLFESGRPTPHAYNHLGSG					
40		600	610	620	630	640	650
	orfl-1	EKDTTKTNGRNLVYQPAEDRTLLLSGGGTNLNGNITQTNGKLFESGRPTPHAYNHLNDH					
		660	670	680	690	700	710
	orfla.pep	WSKMEGIPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKVEGDHLSNHAQAVFGV					
45		660	670	680	690	700	710
	orfl-1	WSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNHAQAVFGV					
		720	730	740	750	760	770
	orfla.pep	APHQSHTICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVXLXXXXXXXXLXGXAXLX					
50		720	730	740	750	760	770
	orfl-1	APHQSHTICTRSDWTGLTNCVEKITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLN					
		780	790	800	810	820	830
	orfla.pep	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSNGNASFNLSNNAQNG					
55		780	790	800	810	820	830
	orfl-1	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNG					
		840	850	860	870	880	890
	orfla.pep	SLTLSDNAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSG					
60		840	850	860	870	880	890
	orfl-1	SLTLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					

-367-

		900	910	920	930	940
	orfla.pep	TELGNLNLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---	LLSVTPPTSVESRFN			
	orfl-1	TELGNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFN				
5		900	910	920	930	940
	orfla.pep	950	960	970	980	990
	orfl-1	950	960	970	980	990
10		960	970	980	990	1000
	orfla.pep	1010	1020	1030	1040	1050
	orfl-1	1010	1020	1030	1040	1050
15		1020	1030	1040	1050	1060
	orfla.pep	1070	1080	1090	1100	1110
	orfl-1	1070	1080	1090	1100	1110
20		1080	1090	1100	1110	1120
	orfla.pep	1130	1140	1150	1160	1170
	orfl-1	1130	1140	1150	1160	1170
25		1140	1150	1160	1170	1180
	orfla.pep	1190	1200	1210	1220	1230
	orfl-1	1190	1200	1210	1220	1230
30		1200	1210	1220	1230	1240
	orfla.pep	1250	1260	1270	1280	1290
	orfl-1	1250	1260	1270	1280	1290
35		1260	1270	1280	1290	1300
	orfla.pep	1310	1320	1330	1340	1350
	orfl-1	1310	1320	1330	1340	1350
40		1320	1330	1340	1350	1360
	orfla.pep	1370	1380	1390	1400	1410
	orfl-1	1370	1380	1390	1400	1410
45		1380	1390	1400	1410	1420
	orfla.pep	1430	1440	1450		
	orfl-1	1430	1440	1450		
50		1440	1450			
	orfla.pep					
	orfl-1					
55						
	orfla.pep					
	orfl-1					
60						

Homology with adhesion and penetration protein hap precursor of *H. influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

	orfl	23	FXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDF	82
	hap	6	FRLNFLTACVSLGASQAWAGHTYFGIDYQYYRDF	65
	orfl	83	KSMKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYNNVDFGAEGXNIXDQXRXTYKIV	142
	hap	66	TSMKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYNDVDFGAEGRN-PDQHRFTYQIV	124

-368-

5
10
15
20

```

orf1 143 KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGR 202
      KRNNY+A + HPY GDYHMPRLHK VT+AEPV MT+ MDG+ Y D+ NYP+RVRIG+GR
hap 125 KRNNYQAWERKHPYDGDYHMPRLHKFVTEAEPVGMTTNMDGKVYADRENYPERVRIGSGR 184

orf1 203 QYWRSEDEPNNRESSYHIA----- 222
      QYWR+D+DE N SSY+++
hap 185 QYWRDDEETNVHSSYYVSGAYRYLTAGNTHQTSGNGNGTVNLSGNVVS PNHYGPLPTG 244

orf1 223 -----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVF 277
      SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF
hap 245 GSKGDSGSPMFIYDAKKQWLINAVLQTGHPPFGRNGFQLIREEFYNEVLAVDTPSVF 304

orf1 278 --YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSSETAREPVYHA 334
      Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A
hap 305 QRYIPPINGHYSFVSNNDGTGKLTLTRPSKDGSKAKSEVGTVKLFNPSLNQTAKHEV-KA 363

orf1 335 AGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGGLYFQGDFTV-SPENNETWQGA 393
      A G N Y+PR+ G+NI D+GKG L + +NINQAGGLYF+G+F V +NN TWQGA
hap 364 AAGYNIYQPRMEYGKNIYLGDKGKTLTIENNINQAGGLYFEGNFVVKGKQNNITWQGA 423

orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT 423
      GV I +D+TV WKV+ NDRLSKIG GTL
hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGTL 453

```

25 Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

30
35
40

```

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTSL 98
      DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS
hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTLIDHSQFTLSNNATQTGNIKLS 792

orf1 99 GNAKANVSHSALNGNVSLADKAVHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158
      +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N
hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTTVTLENATWTMPSTTLQN 852

orf1 159 LNLDNATITLNSAYRHDAAQAQTGSATDAPXXXXXXXXXXXXLXVTPPTSVESRFNTLTVN 218
      L L+N+T+TLNSAY + S+ +AP L T PTS E RFNTLTVN
hap 853 LTLNNSTVTLSAY-----SASSNNAPRHRS-----LETETTP TSAEHRFNTLTVN 899

orf1 219 GKLNQGTFRFMSELFGYRSDKCLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKP 278
      GKL+GQGTFF+ F S LFGY+SDKCLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP
hap 900 GKLSGQGTFOFTSSLFYKSDKCLKLSNDAEGDYTLVRNTGKEPVTLEQLTLIESLDNKP 959

orf1 279 LSENLFNFTLQNEHVDAGA 296
      LS+ L FTL+N+HVDAGA
hap 960 LSDKLKFTLENDHVDAGA 977

```

45 Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

50
55
60
65

```

Orf1 1 LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGFILFSHNR 60
      LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R
hap 1135 LDRLFVDQAQSAVWTNIAQDKRRYDSDAFRAYQQTNLRLQIGVQKALANGRIGAVFHSR 1194

orf1 61 TENTFDDGIGNSARLAHGAVFGQYGIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLYHYG 120
      ++NTFD+ + N A L + F QY K R+ ++YG
hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGDLOFGVNVGTGISASKMAEEQSRKIHRKAINYG 1254

orf1 121 IQARYRAGFGGGFIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA 180
      + A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P
hap 1255 VNASYQFRLGQLGIQPYFGVNRIFYERENYQSEEVVKTPSLAFNRYNAGIRVDYTFPT 1314

orf1 181 QHISITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA 240
      +IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +
hap 1315 DNISVKPYFFVNYVDVSNANVQTTVNLTVLQPPFGRYWQKEVGLKAEILHFQISAFISKS 1374

orf1 241 KGPQLEAQHSAGIKLGYRW 259
      +G QL Q + G+KLGWRW
hap 1375 QGSOLGKQQNVGVKLGWRW 1393

```

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orfl.pep	MKTDDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYRDFAEEN	60
	orflng	MKTDDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYRDFAEEN	60
10	orfl.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGVQYIVSVAHNGGYN	120
	orflng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN	120
15	orfl.pep	NVDFGAEGXNIXDQXRXTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSY	180
	orflng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSY	179
20	orfl.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNRESSYHIAS-----	223
	orflng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orfl.pep	-----GSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	255
	orflng	GGTVNLGSEKIKHSPYGFLLPTGGSGSGSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	289
30	orfl.pep	FQLVRKDWFYDEIFAGDTHSVFYEPONGKYSFNDNNGTGKINAKHEHNSLPNRLKTRT	315
	orflng	FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNGAGKIDAKHKHYSLPYRLKTRT	359
35	orfl.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEKGKELILTSNINQAGAGGLY	375
	orflng	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDKGKELILTSNINQAGAGGLY	422
40	orfl.pep	FQGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orflng	FEGNFTVSPKNNETWQAGGVHISDGTSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
45	orfl.pep	DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orflng	FGVAPHQSHTICTRSDWTGLTSCTEKTITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
50	orfl.pep	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSA	803
	orflng	TFNGNL-VQAE TRTIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
55	orfl.pep	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
	orflng	VQNGSLTSLDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWT	893
60	orfl.pep	LPSGXELGNLNDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRLXVTPPTSVE	923
	orflng	LPSGTELGNLNDNATITLNSAYRHDAAGAQTGSAADAPRRRSRRS---LLSVTPPTSVE	950
65	orfl.pep	SRFNTLTVNGKLNQGTFRFMSELFGYRSDKLAESSEGTYTAVNNTGNPEASLEQLT	983
	orflng	SRFNTLTVNGKLNQGTFRFMSELFGYRSGKLAESSEGTYTAVNNTGNPEVPSLEQLT	1010
70	orfl.pep	VVEGKDNKPLSENLFNLQNEHVDAGAW	1011
	orflng	VVEGKDNTPLSENLFNLQNEHVDAGAWRYQLIRKDGFRHLNPNVKEQELSDKLGKAGET	1070
75	orfl.pep	LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orflng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
80	orfl.pep	AYRQOTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGA VFGQYIGIDRFY	1271
	orflng	AYRQOTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGA VFGQYIGIRFD	1299

-370-

orfl1.pep IGISAGAGFSSGSLSDGIGXKRRRLVHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
|||||
orflng IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1359

5 orfl1.pep RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1391
|||||
orflng RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1419

10 orfl1.pep AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQHSAGIKLGYRW 1440
|||||
orflng AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQHSAGIKLGYRW 1468

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCCTAA
51 AACCGGCCGC ATCCGCTTCT CGCCCGCTTA CTTAGCCATA TGCCGTGTCGT
15 101 TCGGCATTCT GCCCCAAGCC CGGGCGGGAC ACACTTATTT CGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGCGCAAT
251 CGATGACGAA AGCCCCGATG ATTGATTTTT CTGTGGTATC GCGTAACGGC
301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATAACCG
20 351 CGGCTATAAC AATGTTGATT TTGGTGCGGA GGAAGCAAT CCCGATCAGC
401 ACCGCTTTTC TTACCAAAAT GTGAAAAGAA ATAATTATAA AGCAGGGACT
451 AACGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCACAAAT
501 TGTACAGAT GCAAAACCTG TTGAGATGAC CAGTTATATG GATGGGTGGA
551 AATACGCTGA TTTAAATAAA TACCCTGATC GTGTTGCAAT CGGAGCAGGC
25 601 AGACAATATT GCGCGTCTGA TGAAGACGAA CCCAATAACC GCGAAAGTTC
651 ATATCATATT GCAAGCGCAT ATTCTTGGCT CGTCGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG CGAAAAAATT
751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
30 851 ATGGGGTATT GCAAACAGGC AACCCTTATA TAGGAAAAAG CAATGGCTTC
901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTACGTA TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTAAACG
1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAAACATAA AACTATTCT
1051 CTACCTTATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTCTTT
35 1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCTT TATTGACAAA
1201 GGAAAAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAACCGT
1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGATG GCAGTACCGT TACTTGAAA
40 1351 GTAAACGCGG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGACG
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAAGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCCGCTTT CGCGCGGGAC
45 1601 GTTTGGATTG GAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAA CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCGGCAAT AACACAACCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC
50 1851 GGATCGCACT TTAATGCTTT CCGGCGGAAC AAATTAAAC GGCAATATCA
1901 CGCAACAAA CCGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGCACGCC
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCCACAAGG
2001 AGAAATCGTG TGGACAACG ATTGGATCGA CCGCACATTT AAAGCGGAAA
2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCGCAA TGTTGCCAAA
55 2101 GTGGAAGGCG ATTGGCATT T AAGCAATCAC GCCCAAGCAG TTTTCGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGT
2201 TGACAAGTTG TACCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCAGCTCA
60 2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGACGGCG
2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCAAAA CCGCAACCTC
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACACGCGC
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AAACGTAAAGC
2551 CATTCCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
65 2601 TTTTGAAAAC AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCGGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCCTATCG
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TGCGGCAGAT GCGCGCGCGC
2801 GCCGTTTCGCG CCGTCCCTA TTATCCGTTA CGCCGCCAAC TTCGCGAGAA
70 2851 TCCCGTTTCA ACACGCTGAC GGTAAACGGC AAATTGAACG GTCAGGGAAC

-371-

2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC
 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC
 3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGAAG GAAAAGACAA
 3051 CACACCGCTG TCCGAAAATC TTAATTTCAC CCTGCaaaAc gaacacgtcg
 5 3101 atgccggcgc atggCGTTAT CAGCTTATCC gcaaagacgG CGAGTTCCgc
 3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGgc
 3201 gggagaaACA GAggcccgcT TGACGGCAAA ACAGGCacaA CTTGCCGCCA
 3251 AAcacagggc ggaaaAAGAC AACGcgcaaa gccttgAcgc gctgattgcg
 3301 gCggggcgca atgccaccga AAAGGCagaa agtgttgccg aaccgGCCCCG
 10 3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCAGAG GAAGAGAAAA
 3401 AACGGGTGCA GGGGATAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG
 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG
 3501 GGATTTGCCG CAACCGCAGC CCAACCGCA ACCCAACCG CAGCGCGACC
 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC
 15 3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCGAAGA
 3651 CCGCCGCAAC GCCGTTTGA CAAGCGGCAT CCGGGACACC AAACACTACC
 3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAAATC
 3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGGCATCC TGTTTTCGCA
 3801 CAACCGGACC GGAACACCT TCGACGACGG CATCGGCAAC TCGGCAAGCG
 20 3851 TTGCCACGG TGCCGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC
 3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTT CAGACGGCAT
 3951 CAGAGGCAAA ATCCGCCGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT
 4001 ACCGCGCAGG TTTCCGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG
 4051 CGTATTTTCG TCCAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC
 25 4101 CACCCCGGGC CTTGCATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT
 4151 ATTCAATCAA ACCGGCGCAA CACATTTCCA TCACGCCTTA TTTGAGCCTG
 4201 TCCTATACCG ATGCCGCTTC CGGCAAAGTC CGAACGCGCG TCAATACCGC
 4251 CGTATTGGCG CAGGATTTTC GCAAACCCG CAGTGGCGAA TGGGGCGTAA
 4301 ACGCCGAAAT CAAAGGTTTC ACGCTGTCCC TCCACGCTGC CGCCGCCAAG
 30 4351 GGGCCGCAAT TGAAGCGCA GCACAGCGCG GGCATCAAT TAGGCTACCG
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPOA RAGHTYFGIN
 51 YQYYRDFEEN KGFVAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVVSRRNG
 35 101 VAALAGDQYI VSAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT
 151 NGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG
 201 RQYWRSEDEE PNNRESSYHI ASAYSWLVG NTFQNGSGG GTVNLGSEKI
 251 KHSYPGFLPT GGSFGDSGSP MFIYDAQKQK WLINGVLOTG NPYIGKSNGF
 301 QLVRKDWFYD EIFAGDTHSV FYEPHQNGKY FFNDNNNGAG KIDAKHKHYS
 40 351 LPYRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDK
 401 GKGEILITSN INQAGGLYF EGNFTVSPKN NETWQAGVH ISDGSTVTK
 451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQGGKQA
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIONT
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNGWFGEKD
 45 601 ATKTNGLNL NYPFEEADRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
 651 YNHLGSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK
 701 VEGDWHLNHN AQAVFGVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS
 751 LSKTDVGRNV SLADHAHLNL TGLATFNGNL VQAEIRTRIR RANATQNGNL
 801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS
 50 851 HSAIENGVS LADKAVFHFEN SRFTGKISGG KDTALHLKDS EWTLPSTGTEL
 901 GNINLNDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRRSL LSVTPPTSAE
 951 SRFNTLTUNG KLNQGTGTRF MSELFYGRSG KLKLAESSEG TYTLAVNNTG
 1001 NEPVSLQELT VVEGKDNTPL SENLNFLLQN EHVDAWAWRY QLIRKDGFEF
 1051 LHNVPVKEQL SDKLGKAGET EAALTAKQAQ LAAKQQAQEKD NAQSLDALIA
 55 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK DTALAKQREA
 1151 ETRPATTAFF RARRARRDLP QPQPQPQPQP QRDLSRYAN SGLSEFSATL
 1201 NRVFAVQDEL DRVFAEDRRN AVWTSGIRD KHYRSQDFRA YRQQTDLRQI
 1251 GMQKNLGSGR VGILFSHNRT GNTFDDGIGN SARLAHGAVF GQYIGIRFDI
 1301 GISAGAGFSS GSLSDGIRGK IRRRVLYHYGI QARYRAGFGG FGIPEPHIGAT
 60 1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSEKPAQ HISITPYLSL
 1401 SYTDAASGKV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLHAAAAAK
 1451 GPQLEAQHSA GIKLGYRW*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

-372-

		10	20	30	40	50	60
	orfl-1.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN					
	orflng-1						
5		10	20	30	40	50	60
	orfl-1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN					
10	orflng-1						
		70	80	90	100	110	120
	orfl-1.pep	NVDFGAEGSNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSYM					
15	orflng-1						
		130	140	150	160	170	180
	orfl-1.pep	DGRKYIDQNNYPDRVRIGAGRQYWRSEDEPNRESSYHIASAYSWLVGNTFAQNGSGG					
20	orflng-1						
		190	200	210	220	230	240
	orfl-1.pep	GTVNLGSEKIKHSPYGFLLPTGGSGSGSPMFIYDAQKQKWLINGVLQTNPNYIGKSNFG					
25	orflng-1						
		250	260	270	280	290	300
30	orfl-1.pep	QLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHSLPYRLKTRTV					
35	orflng-1						
		310	320	330	340	350	360
	orfl-1.pep	QLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLYF					
40	orflng-1						
		370	380	390	400	410	420
	orfl-1.pep	QGDFTVSPENNETWQAGVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGSIS					
45	orflng-1						
		430	440	450	460	470	480
	orfl-1.pep	VGDGTVILDQQADDQKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGRLDLNGH					
50	orflng-1						
		490	500	510	520	530	540
	orfl-1.pep	SLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITTTGNNNLSKKEIAYNGWFGEKD					
55	orflng-1						
		550	560	570	580	590	600
60	orfl-1.pep	TTKTNGRLNLVYQPAEDRTLLSGGTNLNGNITQTNGKLEFFSGRPTPHAYNHLNDHWSQ					
65	orflng-1						
		610	620	630	640	650	660
	orfl-1.pep	KEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVGDWHLNSHAQAVFGVAPH					
70	orflng-1						
		670	680	690	700	710	720
	orfl-1.pep	MEGIPQGEIVWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLNSHAQAVFGVAPH					
	orflng-1						

-373-

		730	740	750	760	770	780
	orf1-1.pep	QSHTICTRSDWTGLTNCVEKTTITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNL					
	orf1ng-1	QSHTICTRSDWTGLTSCTEKTTITDDKVIASLSKTDIRGNVSLADHAHLNLTGLATLNGNL					
5		730	740	750	760	770	780
	orf1-1.pep	SANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLT					
	orf1ng-1	SAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNAVQNGSLT					
10		790	800	810	820	830	840
	orf1-1.pep	LSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTEL					
	orf1ng-1	LSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWTLPSGTEL					
15		850	860	870	880	890	900
	orf1-1.pep	GNNLNDNATITLNSAYRHDAAQAQTSATDAPRRRSRRSRRSLLSVTPPTSVESEFRNTLT					
	orf1ng-1	GNNLNDNATITLNSAYRHDAAQAQTSAAADAPRRRSR---RLLSVTPPTSASEFRNTLT					
20		910	920	930	940	950	960
	orf1-1.pep	VNGKLNQGTFRFMSELFYGRSDKLKLAESSEGTYYLAVNNTGNEPASLEQLTVVEGKDN					
	orf1ng-1	VNGKLNQGTFRFMSELFYGRSGKLKLAESSEGTYYLAVNNTGNEPVSLQLTVVEGKDN					
25		970	980	990	1000	1010	1020
	orf1-1.pep	KPLSENLFNFTLQNEHVDAGAWRYQLIRKDGFEFRLHNPVKEQELSDKLGKA-----					
	orf1ng-1	TPLSENLFNFTLQNEHVDAGAWRYQLIRKDGFEFRLHNPVKEQELSDKLGKAGETEAALTAK					
30		1030	1040	1050	1060	1070	
	orf1-1.pep	----EAKKQAEKDNAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAEKEEKKRVQ					
	orf1ng-1	QAQLAAKQAEKDNAQSLDALIAAGRDATEKAESVAEPARQAGGENAGIMQAEKEEKKRVQ					
35		1080	1090	1100	1110	1120	
	orf1-1.pep	ADKDTALAKQREAEATRPATTAFPRARRARDLPQLQPQPQPQORDLISRYANSGLSEFS					
	orf1ng-1	ADKDTALAKQREAEATRPATTAFPRARRARDLPQPQPQPQPQORDLISRYANSGLSEFS					
40		1130	1140	1150	1160	1170	1180
	orf1-1.pep	ATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYSQDFRAYRQQTDLRQIGMQKNLG					
	orf1ng-1	ATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYSQDFRAYRQQTDLRQIGMQKNLG					
45		1190	1200	1210	1220	1230	1240
	orf1-1.pep	SGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYIGIDRFYIGISAGAGFSSGSLSDGI					
	orf1ng-1	SGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFGQYIGIRFDIGISAGAGFSSGSLSDGI					
50		1250	1260	1270	1280	1290	1300
	orf1-1.pep	GGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR					
	orf1ng-1	RGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR					
55		1310	1320	1330	1340	1350	1360
	orf1-1.pep	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
	orf1ng-1	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
60		1370	1380	1390	1400	1410	1420
	orf1-1.pep	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
	orf1ng-1	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
65		1380	1390	1400	1410	1420	1430
	orf1-1.pep	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
	orf1ng-1	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
70		1380	1390	1400	1410	1420	1430

5

10

15

20

25

30

35

40

45

50

55

60

65

70

-375-

5 orflng-1.pep 600 610 620 630 640 650
E K D A T K T N G R L N L N Y Q P E E A D R T L L L S G G T N L N G N I T Q T N G K L F F S G R P T P H A Y N H L G S G
p45387 E T D K N K H N G R L N L I Y K P T T E D R T L L L S G G T N L K G D I T Q T K G L F F S G R P T P H A Y N H L N K R
570 580 590 600 610 620

10 orflng-1.pep 660 670 680 690 700 710
W S K M E G I P Q G E I V W D N D W I D R T F K A E N F H I Q G G Q A V V S R N V A K V E G D W H L S N H A Q A V F G V
p45387 W S E M E G I P Q G E I V W D H D W I N R T F K A E N F Q I K G G S A V V S R N V S S I E G N W T V S N N A N A T F G V
630 640 650 660 670 680

15 orflng-1.pep 720 730 740 750 760 770
A P H Q S H T I C T R S D W T G L T S C T E K T I T D D K V I A S L S K T D I R G N V S L A D H A H L N L T G L A T L N
p45387 V P N Q Q N T I C T R S D W T G L T T C Q K V D L T D T K V I N S I P K T Q I N G S I N L T D N A T A N V K G L A K L N
690 700 710 720 730 740

20 orflng-1.pep 780 790 800 810 820 830
G N L S A G G D T H Y T V T R N A T Q N G N L S L V G N A Q A T F N Q A T L N G N T S A S D N A S F N L S N N A V Q N G
p45387 G N V T L ----- T N H S Q F T L S N N A T Q I G
750 760 770

25 orflng-1.pep 840 850 860 870 880 890
S L T L S D N A K A N V S H S A L N G N V S L A D K A V F H F E N S R F T G K I S G G K D T A L H L K D S E W T L P S G
p45387 N I R L S D N S T A T V D N A N L N G N V H L T D S A Q F S L K N S H F S H Q I Q G D K G T T V T L E N A T W T M P S D
780 790 800 810 820 830

30 orflng-1.pep 900 910 920 930 940 950
T E L G N L N L D N A T I T L N S A Y R H D A A G A Q T G S A A D A P R R S R R S L L S V T P P T S A E S R E N T L T
p45387 T T L Q N L T L N N S T I T L N S A Y ----- S A S N N T P R R S --- L E T E T T P T S A E H R E N T L T
840 850 860 870

35 orflng-1.pep 960 970 980 990 1000 1010
V N G K L N G Q G T F R F M S E L F G Y R S G K L K L A E S S E G T Y T L A V N N T G N E P V S L E Q L T V V E G K D N
p45387 V N G K L S G Q G T F Q F T S S L F G Y K S D K L K L S N D A E G D Y I L S V R N T G K E P E T L E Q L T L V E S K D N
880 890 900 910 920 930

40 orflng-1.pep 1020 1030 1040 1050 1060 1070
T P L S E N L N F T L Q N E H V D A G A W R Y Q L I R K D G E F R L H N P V K E Q E L S D K L G K A G E T E A A L T A K
p45387 Q P L S D K L K F T L E N D H V D A G A L R Y K L V K N D G E F R L H N P I K E Q E L H N D L V R A E Q A E R T L E A K
940 950 960 970 980 990

45 orflng-1.pep 1080 1090 1100 1110 1120 1130
Q A Q L A A Q Q A E K D N A Q S L D A L I A A G R N A T - E K A E S V A E P A R Q A G G E N A G I M Q A E E E K K R V
p45387 Q V E P T A K T Q T G E P K V R S R R A A R A F P D T L P D Q S L L N A L E A K Q A E - L T A E T Q K S A K A T K K V
1000 1010 1020 1030 1040 1050

50 orflng-1.pep 1140 1150 1160 1170 1180 1190
Q A D K --- D T A L A K Q R E A E T R P A T T A F P R R A R R A R D - L P Q P Q P Q P Q P Q R D L I S R Y A N S G
p45387 R S K R A V F S D P L L D Q S L F A L E A A L E V I D A P Q Q S E K D R L A Q E E A E K Q - R K Q K D L I S R Y S N S A
1060 1070 1080 1090 1100 1110

55 orflng-1.pep 1200 1210 1220 1230 1240 1250
L S E F S A T L N S V F A V Q D E L D R V F A E D R R N A V W T S G I R D T K H Y R S Q D F R A Y R Q Q - T D L R Q I G
p45387 L S E L S A T V N S M L S V Q D E L D R L F V D Q A Q S A V W T N I A Q D K R R Y D S D A F R A Y Q Q Q K T N L R Q I G
1120 1130 1140 1150 1160 1170

60 orflng-1.pep 1260 1270 1280 1290 1300 1310
M Q K N L G S G R V G I L F S H N R T G N T F D D G I G N S A R L A H G A V F G Q Y G I G R F D I G I S A G A G F S S G
p45387 V Q K A L A N G R I G A V F S H S R S D N T F D E Q V K N H A T L T M M S G F A Q Y Q W G D L Q F G V N V G T G I S A S
1180 1190 1200 1210 1220 1230

65 orflng-1.pep 1260 1270 1280 1290 1300 1310
M Q K N L G S G R V G I L F S H N R T G N T F D D G I G N S A R L A H G A V F G Q Y G I G R F D I G I S A G A G F S S G
p45387 V Q K A L A N G R I G A V F S H S R S D N T F D E Q V K N H A T L T M M S G F A Q Y Q W G D L Q F G V N V G T G I S A S
1180 1190 1200 1210 1220 1230

70 orflng-1.pep 1260 1270 1280 1290 1300 1310
M Q K N L G S G R V G I L F S H N R T G N T F D D G I G N S A R L A H G A V F G Q Y G I G R F D I G I S A G A G F S S G
p45387 V Q K A L A N G R I G A V F S H S R S D N T F D E Q V K N H A T L T M M S G F A Q Y Q W G D L Q F G V N V G T G I S A S
1180 1190 1200 1210 1220 1230

		1320	1330	1340	1350	1360	1370
orf1ng-1.pep		SLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGL					
		::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::					
p45387		KMAEEQSRKIHRRKAINYGVNASYQFRLGQLGIQPYFGVNRYPFIERENYQSEEVRVKTPSL					
		1240	1250	1260	1270	1280	1290
		1380	1390	1400	1410	1420	1430
orf1ng-1.pep		AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEW					
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::					
p45387		AFNRYNAGIRVDYTFPTDNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEV					
		1300	1310	1320	1330	1340	1350
		1440	1450	1460	1469		
orf1ng-1.pep		GVNAEIKGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX					
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::					
p45387		GLKAEILHFQISAFISKSQGSQGLGKQQNVGKLGYSRW					
		1360	1370	1380	1390		

20 Example 78

25
30

```
1      . . AAGGTGTGGC AATTGTGCGA AGA . CCGCTG CGTGCCGTCG TGCCTGCCGA
51     CAGTTTTTGAA CCGACCGCGC AAAAATTGAA CCTGTTTAAG GCGGGTGCGG
101    CAACCATTTT GTTTTATGAA GATCAAAATG TCGTCAAAGG TTTGCAGGAG
151    CAGTTCCTGT CTTATGCCGC TAACCTCCCG GTTTGGCGGg ATCAGGCAA
201    CGCGATGGTG CAGTATGCCG TTGGACGAC ACTTGGCCGC GTCGGCGTAG
251    GTGCAAACCT GCAACATTAC AATCCCTTGC CCGATGCGGC GATTGCCAAA
301    GCGTGGAATA TCCCCGAAAA CTGGTTGTTG CGCGCACAAA TGTTTATCGG
351    CCGTATTGAA GGGCGGCAG GTGAAAAGAC CTTTGAACCC GTTGCAAGAC
401    GTTTGAAAGT GTTCGGCGCA TAA
```

```

      1  ..KVWQFVEXPL RAVVPADSFE PTAQKLNLFK AGAATILFYE DQNVVKGLQE
     51  QFPAYAANFP VWADQANAMV QYAVWTTLAA VGVGANLQHY NPLPDAAIK
    101  AWNIPENWLL RAQMVIIGGIE GAAGEKTFEP VAERLKVFGA *

```

40

1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAATT
51	GAACCTGTTT	AAGCGGGTG	CGCAACCAT	TTTGTTTTAT	GAAGATCAAA
101	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATCG	CGCTAACTTC
151	CCCGTTTGCG	CGGATCAGGC	AAACCGCGATG	GTGCGAGTATG	CCGTTTGGAC
201	GACACTTGCC	GCGGTGCGCG	TAGGTGCAAA	CCTGCAACAT	TACAATCCCT
251	TGCCCCGATG	GCGGATTGCC	AAAGCGTGGA	ATATCCCCGA	AAACTGGTTG
301	TGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG	CAGGTGAAAA
351	GACCTTTGAA	CCCGTTTGAG	AACGTTTGAA	AGTGTTCGGC	GCATAA

```

45      1  ..LRADVPSDF EPTAQLNLF KAGAATILFY EDQNVVKGLQ EQFPAYAAAF
      51  PVWADQANAM VQYAVWTTLA AVGVGANLQH YNPLDAAIA KAWNIPENWL
     101  LRAQMVIGGI EGAAGEKTFE PVAERLKVFG A*

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 • 20 30

-377-

```

      orf6.pep                                KVVQFVEXPLRAVVPADSFEPKQLNLFK
      orf6a      QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPKQLNLFK
                    40      50      60      70      80      90
5
      orf6.pep      AGAATILFYEDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
      orf6a      AGAATILFYEDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
10
                    100      110      120      130      140      150
      orf6.pep      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
      orf6a      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
15
                    160      170      180      190      200

```

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

```

1  ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
51  TTCGTAAAT AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG
20 101  TCGAACACGC CGTTTTGCAC ACACCTTCTT CGTTCAATTC CCAATCTGCC
151  CGTGTGGTCG TGCTGTTTGG CGAAGAGCAT GATAAGGTGT GGCAATTTGT
201  CGAAGACGCG CTGCGTGCCG TCGTGCTGCG CGACAGTTTT GAACCGACCG
251  CGCAAAAATT GAACCTGTTT AAGGCGGGTG CGGCAACTAT TTTGTTTTAT
301  GAAGATCAAA ATGTCGTCAA AGGTTTGCAG GAGCAGTTCC CTGCTTATGC
25 351  CGCCAACCTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG
401  CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT
451  TACAATCCCT TGCCCGATGC GCGGATTGCC AAAGCGTGGA ATATCCCCGA
501  AAATGTTTGT TTGCGCGCAC AAATGTTTAT CGGCGGTATT GAAGGGGCGG
551  CAGGTGAAAA GACCTTTGAA CCAGTTGCAG AACGTTTGAA AGTGTTCCGG
30 601  GCATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

```

1  MTRQSLQQA ESRRSIYSLN KNLPVGKDEI VQIVEHAVLH TPSSFNSQSA
51  RVVVLFGEEH DKVVQFVEDA LRAVVPADSF EPTAQKLNLF KAGAATILFY
35 101  EDQNVVKGLQ EQFPAYAAFP PVWADQANAM VQYAVWTTLA AVGVGANLQH
151  YNPLPDAAIA KAWNIPENWL LRAQMVIIGI EGAAGEKTFE PVAERLKVFG
201  A*

```

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

```

40      orf6a.pep      TPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPKQLNLFKAGAATILFY
      orf6-1      LRAVVPADSFEPKQLNLFKAGAATILFY
                    10      20      30
45      orf6a.pep      EDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
      orf6-1      EDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
50
                    40      50      60      70      80      90
      orf6a.pep      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
      orf6-1      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
55
                    100      110      120      130

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

```

      orf6.pep                                KVVQFVEXPLRAVVPADSFEPTAQKLNLFK    30
      orf6ng      SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLFK    64
5      orf6.pep      AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY    90
      orf6ng      AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGAGANLQHY    124
10     orf6.pep      NPLPDAAIAKAWNIPENWLLRAQMVIGGIEGAAGEKTFFPVAERLKVFGA    140
      orf6ng      NPLPDVAIAKAWNIPENWLLRAQMVIGGIEGAAGEKVFFPVAERLKVFGA    174

```

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

```

15      1  ATGGCCGTTG  CGTCAAATGT  CAGCTTGGAT  ATGTCCAATC  CTACGGTGTG
      51  ACGCATGGGA  TTACCCTTAT  ATATTGCGTC  CCTAAGAAGG  GGC GCAATAT
      101  ATAAGGTGTG  GCAATTTGTC  GAAGACGCGC  TGCCTGCCGT  CGTGCCGTGCC
      151  GACAGTTTTG  AACCGACCGC  GCAAAAATTG  AAGCTGTTTA  AGGCGGGCGC
      201  GGCAACCATT  TTGTTTTATG  AAGATCAAAA  TGTCGTCAAA  GGTTCGAGG
      251  AGCAGTTCCT  TGCTTATGCC  GCCAACTTTC  CCGTTGGGCG  GGACCAGGCG
20      301  AACGCTATGG  TACAGTATGC  CGTCTGGACG  ACGTTCGCCG  CGGTCGGTGC
      351  AGGTGCAAAT  CTGCAACATT  ACAACCCCTT  GCCCGATGTG  GCGATTGCTA
      401  AAGCGTGGAA  TATTCCCGAA  AACTGGCTGT  TGCGCGCGCA  AATGGTTATC
      451  GGTGGTATTG  AAGGGGcggc  aggtgaaaaa  gtctttgaac  CCGTTGCgga
      501  acgtttgAAA  GTGTTGCGCG  CATAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

```

30      1  MAVASNVSLD  MSNPTVLRMG  LPLYIASLRR  GAIYKVVQFV  EDALRAVVPA
      51  DSFEPTAQL  KLFKAGAATI  LFYEDQNVVK  GLQEQFPAYA  ANFPVWADQA
      101  NAMVQYAVWT  TLAAVGAGAN  LQHYNPLPDV  AIAKAWNIPE  NWLLRAQMVI
      151  GGIEGAAGEK  VFEPVAERLK  VFGA*

```

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

```

      orf6-1.pep                                10      20      30
      LRAVVPADSFEPTAQKLNLFKAGAATILFY
35     orf6ng      PTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLFKAGAATILFY
      20      30      40      50      60      70

      orf6-1.pep      40      50      60      70      80      90
      EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
40     orf6ng      EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGAGANLQHYNPLPDVAIA
      80      90      100     110     120     130

      orf6-1.pep      100     110     120     130
      KAWNIPENWLLRAQMVIGGIEGAAGEKTFFPVAERLKVFGAX
45     orf6ng      KAWNIPENWLLRAQMVIGGIEGAAGEKVFFPVAERLKVFGAX
      140     150     160     170

```

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

```

55      1  ..GGCTACAAC  ACCTGTTGCG  GCGCGGCAGC  CGCATCGCCA  ACTACCAAAT
      51  CAACGGCATC  CCCGTTGCCG  ACGCGCTGGC  CGATACGGGt  CAATGCCAAC
      101  ACCGCGCCT  ATGAGCGCGT  AGAAGTCGTG  CGCGGCGTGG  CGGGGCTGCT
      151  GGACGGCACG  GCGGAGCCTT  CCGCCACCGT  CAATCTGGTG  CGCAAACGCC
      201  TGACCCGCAA  GCCATTGTTT  GAAGTCCGCG  CCGAAGCgGG  CAACCGcAAA

```


-379-

5
251 CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crc
301 rCTGCGCGGC CGCCTGGTTT CCAcCTTCGG AC GCGGCGGAC TCGTGGCGGC
351 GGCGCGAAGC GAGCCGskAT GCCGAATCT AC GGCATTTT GGAATACGAC
401 ATCGCACCGC AAACCCGCGT CCACGCArGC ATGGACTACC AGCAGGCGAA
451 AGAAACCGCC GACGCGCCGC TCAGcTACGC CGTGTACGAC AGCCAAGGTT
501 ATGCCACCGC CTTCGGCCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC
551 AGCCACCACC GTGCGCTCAA CCTGTTCGCC GGCATCGAAC ACCGCTTCAA
601 CCAAGACTGG AAACCTCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10
1 ..GYNLYFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL
51 DGTGEP SATV NLVRKRLTRK PLFEVRAEAG NRKHFG LDAD VSGSLNTEXX
101 LRGR LVSTFG RGDSWRRRER SRXAE LYGI EYDIAPQTRV HAXMDYQQA
151 ETADAPLSYA VYDSQGYATA FGPKN DPATN WANS HHRLN LFAGIEHRFN
201 QDWKLKAEYD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
51 CGCGCAGGCC GATGTTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA
101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC
151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGTCGGGC TGCCCATGAC
201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GCGGACCGGC
301 ACCAGCCGCC AGATTTACGG CTCGACCGC GCGGGCTACA ACTACCTGTT
351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCGGTTG
401 CCGACGCGCT GGCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
251 GTAGAAAGTCG TGCGCGGCGT GGCGGGGCTG CTGGACGGCA CCGGCGAGCC
501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCTGACCCGC AAGCCATTGT
551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCCG GCTGGACCGC
601 GAGGTATCGG GCAGCCTGAA CACCGAAGGC ACGTGC GCG GCGCGCTGGT
651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG
301 701 ATGCCGAAT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC
751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCGCCA CCGTGC GCTC
901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
351 951 AGCCGAATAC GACTACACCC GCAGCGGCTT CCGCCAGCCC TACGGCGTAG
1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCACCGA CCTGATTCCC
1051 GGTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTATTGAT
1101 CGGCAAATAC CGCCTGTTTC GCGCGAACA CGATTAAATC GCGGGTATCA
1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCGC CATCATCCCC
401 1201 AAGCCCATTC CCAACGCCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGGCGGCAAA
1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCGGACAA CCTTTCGCTG
1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACGCG
441 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCCTACACAG
1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC
1501 AGCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG
1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
501 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
1701 CGCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCGCA
1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT
1851 CAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGCC
551 1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCGCGCGCG CCGACAACAG
2001 CCGCCAAAAA GCCTACGCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCGC CGAACTGTCT CTGAACGTGG ACAATCTGTT CAACAAACAC
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCACTGC GGACAGTGAA
2151 CGCGCGGTTT ACCTATCGGT TTAAATAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

65
1 MTRFKYSLLF AALLPVYAQA DVSVD DPKP QESTELPTIT VTADRTASSN
51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
101 TSRQIYSDR AGYNLYFARG SRIANYQING IPVADALADT GNANTAAYER
151 VEVVRGVAGL LDGTGEP SAT VNLVRKRLTR KPLFEVRAEA GNRKHFG LDA
201 DVSGSLNTEG TLRGR LVSTF GRGDSWRRRE RSRDAELYGI LEYDIAPQTR

251	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWANSRHRAL
301	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQFP	YGVAGVLSID	HNTAATDLIP
351	GYWHADPRTH	SASVSLIGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP
401	NAIPNAYEFS	RTGAYPQPAS	FAQTIPQYGT	RRQIGGYLAT	RFRADNLSL
451	ILGGRYTRYR	TGSYDSRTQG	MTYVSNRFT	PYTGIVFDLT	GNLSLYGSYS
501	SLFVPQSQKD	EHGSYLKPV	GNNLEAGIKG	EWLEGRNLAS	AAVYRARKNN
551	LATAAGRDPS	GNTYYRAANQ	AKTHGWEIEV	GGRTPEWQI	QAGYSQSKTR
601	DQDGSRLNPD	SVPERSFKLF	TAYHFAPEAP	SGWTIGAGVR	WQSEHTDPA
651	TLRIPNPAAK	ARAADNSRQK	AYAVADIMAR	YRFNPRAELS	LNVDNLFNKH
701	YRTQPDHSY	GALRTVNAAF	TYRFK*		

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

15	Orf23	6	FARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRK	65
	PupB	215	WSRGFAIQNYEVDGVPSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK	273
20	Orf23	66	RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFXXXXXXXXXXXXXXAE	125
	PupB	274	RPTAEAQASITGEAGNWDYGTGFVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL	333
25	Orf23	126	LYGILEYDIAPQTRVHAXMDYQOAKETADAPLSYAYVD--SQGYATAFGPKDNPATNWAN	183
	PupB	334	MYGITEFDLSEDTLTVGFSY--LRSDIDSPLRSGPLTRFSTGERTNLKRSLSNAAPDWSY	391
	Orf23	184	SHHRALNLFAGIEHRFNQDWKLKAE	208
	PupB	392	NDHEQTSFSTSIEQQLGNWWSGKIE	416

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N. meningitidis*:

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

-381-

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCAAAACCG CAGGAAAGCA
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC
 5 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
 251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCCTGTTGCA GCGGACCGGC
 301 ACCAGCCGCC AGATTTACGG CTCGACCGCG GCGGGCTACA ACTACCTGTT
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCCG CTATGAGCGC
 10 451 GTAGAAAGTCG TCGCGGGCGT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCCGACCCGC AAGCCATTGT
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCCG GCTGGGCGCG
 601 GACGTATCGG GCAGCCTGAA TGCCGAAGGC ACGCTGCGCG GCCGCTGCT
 651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCAGCGCGAA CGCAGCCCG
 15 701 ATGCCGAATC TGGGAATACG ACATCGCACC GCAAACCTCG
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
 801 GCTCAGCTAC GCGGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
 851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCCGCCA CCGTGCGCTC
 901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAGACT GGAAACTCAA
 20 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
 1051 GGTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCTTAAT
 1101 CGGCAAATAC CGCTGTTTCG GCGCGAACA CGATTTAATC GCGGGTATCA
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC
 25 1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
 1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGCGGCAAA
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
 1351 ATACTCGGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCTACACAG
 30 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CGCTTTACGG CTGCTACAGC
 1501 AGCCTGTTTC TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG
 1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
 35 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
 1751 TCACGCCC GA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCCG
 1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT
 1851 CAACTCTTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGCTGGA
 1901 CATTCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGCC
 40 1951 ACGTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCGC CGAACTGTTC CTGAACGTGG ACAATCTGTT CAACAAACAC
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
 2151 CGCGGCGTTT ACCTATGGGT TTAAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QMRDQNIKA LDRALLQATG
 101 TSRQIYGS DR AGYNYLFARG SRIANYQING IPVADALADT GNANTAAYER
 151 VEVVRGVAGL LDGTGEP SAT VNLVRKRPRTR KPLFEVRAEA GNRKHFGLGA
 50 201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRFRQ YGVAGVLSID HNTAATDLIP
 351 GYWHADPRTH SASVSLIGKY RLFGRHDLI AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPOPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
 55 451 ILGGRYSR YR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNLNAS AAVYRARKNN
 551 LATAAGRDP S GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSEHTDPA
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNDFLNKH
 60 701 YRTQPD RSHY GALRTVNAAF TYRFK*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

		10	20	30	40	50	60
orf23a.pep		MTRFKYSLLF	AALLPVYAQA	DVSVDPPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
65	orf23-1	MTRFKYSLLF	AALLPVYAQA	DVSVDPPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
		10	20	30	40	50	60

-382-

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYL	FARG				
	orf23-1	PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRLALLQATGTSRQIYGSDRAGYNYL	FARG				
5		70	80	90	100	110	120
	orf23a.pep	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP	SATVNLVRKR	PTR			
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP	SATVNLVRKRL	TR			
		130	140	150	160	170	180
	orf23a.pep	KPLFEVRAEAGNRKHFGLDADVSGSLNAEGTLRGRLVSTFGRGDSWRQ	RERSRDAEL	YGI			
15	orf23-1	KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRLVSTFGRGDSWR	RERSRDAEL	YGI			
		190	200	210	220	230	240
	orf23a.pep	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDN	PATNWANSR	HRAL			
20	orf23-1	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDN	PATNWANSR	HRAL			
		250	260	270	280	290	300
	orf23a.pep	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHN	TAATDLIPG	YWHADPR	TH		
25	orf23-1	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHN	TAATDLIPG	YWHADPR	TH		
		310	320	330	340	350	360
	orf23a.pep	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIP	NAYEFSRT	GAYPQP	PAS		
30	orf23-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIP	NAYEFSRT	GAYPQP	PAS		
		370	380	390	400	410	420
	orf23a.pep	FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYSTRYRTG	SYDSRTQ	GMTYV	SANRFT		
35	orf23-1	FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYSTRYRTG	SYDSRTQ	GMTYV	SANRFT		
		430	440	450	460	470	480
	orf23a.pep	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGN	NLEAGIK	GEWLE	GRNLNAS		
40	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGN	NLEAGIK	GEWLE	GRNLNAS		
		490	500	510	520	530	540
	orf23a.pep	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGG	RITPEWQ	IQAGYS	QSKTR		
45	orf23-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGG	RITPEWQ	IQAGYS	QSKTR		
		550	560	570	580	590	600
	orf23a.pep	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQ	SETH	TDPATL	RIPNPA	AK	
50	orf23-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQ	SETH	TDPATL	RIPNPA	AK	
		610	620	630	640	650	660
	orf23a.pep	ARAADNSRQKAYAVADIMARYRFPRAELSLNVDNLFNKHYRT	QPD	RHSY	GALRT	VNAAF	
55	orf23-1	ARAADNSRQKAYAVADIMARYRFPRAELSLNVDNLFNKHYRT	QPD	RHSY	GALRT	VNAAF	
		670	680	690	700	710	720
	orf23a.pep	TYR	FKX				
60	orf23-1	TYR	FKX				
65	orf23a.pep	TYR	FKX				
70	orf23-1	TYR	FKX				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N.gonorrhoeae*:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPG YNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLPD	60
10	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLVSTFGR	120
15	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRLERSRDAELYGILEYDIAPQTRVHAGMDYQAKETADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWNANSHHRLNLFAGIEHFRNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNRNRLNLFAGIEHFRNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising

20 amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTKRP	LFEVRAEAGN	RKHFGGLADV
	101	SGSLNAEGTL	RGRLVSTFGR	GDSWRLERS	RDAELYGILE	YDIAPQTRVH
25	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSNRNRLNL
	201	FAGIEHFRNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGRSIIIPNA
	301	IPNAYEFSRT	GAYPQPSSFA	QTIPOYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGRYSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
30	401	FVPQLQKDEH	GSYLKPVGTN	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
	451	TAAGRDQSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQA	GYSQSKPRDQ
	501	DGSRLNPDSV	PERSFKLFTA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTLSLN	VDNLFNKHYR
	601	TQPDHRSYGA	LRTVNAAFTY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTTCCAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCCATGAC
40	201	CCTGCGCGAA	ATCCCCGAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
	251	GCGACCAAAA	CATCAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACCTACCA	AATCAACGGC	ATCCCCGTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
45	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGACGGCA	CGGGCGAGCC
	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAAC	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCCGC
50	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCTGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
55	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCcgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAAATAC	CgcctGTTTCG	GCCGCGAGCA	CGATTTAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCGC	CATCATTTCC
	1201	AACGCCATTTC	CCAACGCCTA	CGAATTTTCC	CGCACGGGCG	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGCGGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGCTG
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGCAGCT	ACAACAGCCG

-384-

1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG
 1451 GCATCGTGTT CGATCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC
 1501 AGCCTGTTTCG TCCCAGCAATT GCAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGA CATCAAAGGC GAATGGCTTG
 1601 AAGGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
 1651 CTCGCCACCG CAGCAGGACG CGACCAGAGC GGCAACACCT ACTATCGCGC
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGCGCCGCA
 1751 TCACGCCCCG ATGGCAGATA CAGGCAGGCT ACAGCCAAAG CAAACCCCGC
 1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT
 1851 CAAACTCTTC ACCGCCTACC ACTTAGCCCC CGAAGCCCCC AGCGGCCGGA
 1901 CCATcggtGC GGGTGTGCGC CGGCAGGGCG AAACCCACAC CGACCCAGCC
 1951 GCGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG TCGCCAACAG
 2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCAC CGAACTGTCG CTGAACGTGG ACAACCTGTT CAACAAACAC
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1 MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QMRDQNIKT LDRALLQATG
 101 TSROIYGS DR AGNYL FARG SRIANYQING IPVADALADT GNANTAAYER
 151 VEVVRGVAGL PDGTGEP SAT VNLVRKH PTR KPLFEVRAEA GNRKHFGLGA
 201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSNRNAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HSTAATDLIP
 351 GYWHADPRTH SASMSLTGKY RLFGRHDLI AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPQSS FAQTIPQYDT RQIGGYLAT RFRAADNLSL
 451 ILGGYRSYR AGSYNSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFVPQLQKD EHGSYLKPV T GNNLEADIKG EWLEGRNLNAS AAVYRARKNN
 551 LATAAGRDQS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKPR
 601 DQDGSRLNPD SVPERSEKLF TAYHLAPEAP SGRITIGAGVR RQGETHTDPA
 651 ALRIPNPAK ARAVANSRQK AYAVADIMAR YRFNPRTELS LNVNDLNFKNH
 701 YRTQDRHSY GALRTVNAAF TYRFK*

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

35	orf23-1.pep	10 20 30 40 50 60	MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN DGYTVSGTHT
	orf23ng-1	10 20 30 40 50 60	MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN DGYTVSGTHT
40	orf23-1.pep	70 80 90 100 110 120	PLGLPMTLRE IPQSVSVITS QMRDQNIKT LDRALLQATG TSROIYGS DRAGNYL FARG
	orf23ng-1	70 80 90 100 110 120	PFGLPMTLRE IPQSVSVITS QMRDQNIKT LDRALLQATG TSROIYGS DRAGNYL FARG
45	orf23-1.pep	130 140 150 160 170 180	SRIANYQING IPVADALADT GNANTAAYER VEVVRGVAGL PDGTGEP SATVNLVRKH PTR
	orf23ng-1	130 140 150 160 170 180	SRIANYQING IPVADALADT GNANTAAYER VEVVRGVAGL PDGTGEP SATVNLVRKH PTR
50	orf23-1.pep	190 200 210 220 230 240	KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRVSTFGRGDSWRRRERSRDAELYGI
	orf23ng-1	190 200 210 220 230 240	KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRVSTFGRGDSWRQLERSRDAELYGI
55	orf23-1.pep	250 260 270 280 290 300	LEYDIAPQTR VHAGMDYQQA KETADAPLSY AVYDSQGYATAFGPKDNPATNWNANSRHRAL
	orf23ng-1	250 260 270 280 290 300	LEYDIAPQTR VHAGMDYQQA KETADAPLSY AVYDSQGYATAFGPKDNPATNWSNSNRNAL
60	orf23-1.pep	310 320 330 340 350 360	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHN TAATDLIPGYWHADPRTH
	orf23ng-1	310 320 330 340 350 360	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHN TAATDLIPGYWHADPRTH

-385-

		310	320	330	340	350	360
5	orf23-1.pep	370	380	390	400	410	420
	orf23ng-1	370	380	390	400	410	420
10	orf23-1.pep	430	440	450	460	470	480
	orf23ng-1	430	440	450	460	470	480
15	orf23-1.pep	490	500	510	520	530	540
	orf23ng-1	490	500	510	520	530	540
20	orf23-1.pep	550	560	570	580	590	600
	orf23ng-1	550	560	570	580	590	600
25	orf23-1.pep	610	620	630	640	650	660
	orf23ng-1	610	620	630	640	650	660
30	orf23-1.pep	670	680	690	700	710	720
	orf23ng-1	670	680	690	700	710	720
35	orf23-1.pep	TYRFX					
	orf23ng-1	TYRFX					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
50	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPOSVSVITSQQMRDQNIKTLDRL 95 T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL + Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTTSAGTKMQMTQRDIPOSVTIVSQRMEDQQQLQTLGEVM 102
60	Query: 96 LQATGTSRQIYGS DRAGYNILFARGSR IANYQINGIP-----VADALADTGNANTAA 147 G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLGISKSQADSDRALY---YSRGFQIDNYMVDGIPTYFESRWNLGDALS DM-----AL 154
65	Query: 148 YERVEVVRGVAGLPDGTGEP SATVNLVRKHPTRKPLF-EVRAEAGNRKHFG LGADVSGSL 206 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSA AINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSP L 214
70	Query: 207 NAEGLRGLRVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQA KETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNND SWLD RYNSEKTF FSGIVDADLGDLTLSAGYEQRIDVNSPT 274
	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSNRNALNLFAGIEHRFNQDWKLKAEYDYTRSR 326

-386-

```

      +++ G + ++      + A +W+ +      +F ++ +F W+      ++
Sbjct: 275 WGGLEPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334

Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHTSA-SMSLTGKYRLFG 374
5      F + Y A V D      ++ PG+      W++ R A + G Y LFG
Sbjct: 335 FDSKMMYVDAYVNKADGMLVGYPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394

Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432
      R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
10 Sbjct: 395 RQHNLMEFG-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451

Query: 433 QIGGYLATRFRAADNLSLILGGYRSRYRAGSYNSRTQGMTY-VSANRFTPYTGIVFDXXX 491
      Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
15 Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504

Query: 492 XXXXXXXXXXXXFPVQLQKDEHGSYLKPVGTNNLEADIKGEWLEGRLNASAAVYRARKNNL 551
      F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
20 Sbjct: 505 NWSTYASYTSIFQPQNDRDSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564

Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608
      A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
25 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624

Query: 609 PDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAAKARAVANSR 668
      P ++P + K+FT+Y L P P T+G GV Q +TD P RA
30 Sbjct: 625 P-NLPRTTVKMFETSYRL-PVMEPE-LTVGGGVNWQNRVYTDTV-----TPYGTFR-----E 672

Query: 669 QKAYAVADIMARYRFNPRTELSLNVDNLFNKHYRTQPD RH-SYGALRTVNAAFTYRF 724
      Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+P
30 Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTPRNFSITGTYQF 729

```

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described
 35 above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

40 Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

```

      1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
      51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
45 101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
      151 AGCGTCAGcA CGCCTGCTTC GGCGgcGgCa ATCATACCTT CGTCTTCGGA
      201 AACGGGGATA AACGcGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
      251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
      301 CCGTGCGTAC CGCAGACGCT CAAGCCCAT TnTTCAAGAA TGCGTGCCAC
      351 TnAGTCGCCG ACGGGG..

```

50 This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

```

      1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
      51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
      101 PCVPQTLKPI XSRMRATXSP TG..

```


-387-

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

      1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
     51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTCGCCG GGAACGGCAA
    101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
    151 AGCGTCAGCA CGCCTGCTTC GCGGCGGGCA ATCATACCTT CGTCTTCGGA
    201 AACGGGGATA AACGCGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
    251 TGCCGCCTTT TTTACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
    301 CCGTGCGTAC CGCAGACGCT CAAGCCCATT TCTTCAAGAA TGCCTGCCAC
    351 TGAGTCGCCG ACGGCGGGGG TCGGCGCCAG CGACAAGTCG AGAATACCAA
    401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCACG
    451 CGGGTAATTT TGAAAGCAGT TTTCTTCACT ACTTCCGCAA CTTCGGTCAA
    501 TGTCGTTGCA TCTGAATTTT CCAACGCGGC TTTACGACA CCTGGGCGCG
    551 ATACGCCGAC ATTGATAACG GCATCCGCTT CGCCCGAACC ATGAAACGCG
    601 CCCGCCATAA ACGGGTTGTC TTCCACCGCG TTGCAGAACA CGACAATTTT
    651 AGCGCAGCCG AAACCTTCGG GCGTGATTTT CGCCGTGCGT TTGACGGTTT
    701 CGCCCGCCAG CTTGACCGCA TCCATATTGA TACCGGCACG CGTACTGCCG
    751 ATATTGATGG AGCTGCACAC AATATCGGTA GTCTTCATCG CTTCGGGAAT
    801 GGAGCGGATT AACACCTCAT CCGAAGGCGA CATCCCTTTT TGCACCAACG
    851 CGGAAAAACC GCCGATAAAA GACACACCGA TGGCTTTGGC AGCTTTATCC
    901 AAAGTTTGCG CCACGCTGAC GTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

      1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTQ TAVMASSLS
     51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
    101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT
    151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP*NA
    201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LTVSPASLTA SILIPARVLP
    251 ILMELHTISV VFIAAGMERI NTSSEGDIPF CTNAEKPPIK DTPMALAALS
    301 KVCATLT*
  
```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
    35 orf24a.pep MRTAVVLLLIMPMAASSAMMPEMVCAGVSPGTAIISXPTEQTAVIASLNSVSTPASAAA
       orf24      MRTAVVLLLIMPMAASSAMMPEMVCAGVSPGTAIISKPTQ TAVMASSLSVSTPASAAA
      10      20      30      40      50      60
    40 orf24a.pep IIPSSSXTGINAPLKPPTALEAIMPFFFTASFSNAKAAVVPCVPQTLKPISSMRATESP
       orf24      IIPSSSETGINAPLKPPTALEAIMPFFFTASFSNAKAAVVPCVPQTLKPISSMRATESP
      70      80      90      100     110     120
    45 orf24a.pep TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
       orf24      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      130     140     150     160     170     180
    50 orf24a.pep PGPDTPTLITASASPEPXNAPAI XGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
       orf24      PGPDTPTLITASASPEPXNAPAI NGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
      190     200     210     220     230     240
    55 orf24a.pep SILIPARVLPILMELHTISVVFIAAGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
       orf24      SILIPARVLPILMELHTISVVFIAAGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS
      250     260     270     280     290     300
    60 orf24a.pep
       orf24
  
```

-388-

```
orf24a.pep  KVCATLTX
            |||||
orf24       KVCATLTX
```

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```
1  ATGCGCACGG CAGTGGT TTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA
101 TCATATCCAA NCCGACCGAA CAAACGGCGG TCATCGCTTC GAGTTTATCC
151 AACGTCAGCA CGCCTGCTTC GCGCGCGGCA ATCATACCTT CGTCTTCGGA
10 201 NACGGGGATA AACCGGCCAC TCAAACCGCC AACCGCGCTC GAAGCCATCA
251 TGCCGCCCTT TTTACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TCGCGGCCAC
351 CGAGTCGCCG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGATATT CAGCATTTT GAGGCTTCGC GGCCGATGAG TTCGCCACG
15 451 CGGGTAATTT TGAAGGCGGT TTTCTTCACA ACTTCGGCAA CTTCGGTCAA
501 TGTCGTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCCGG
551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACGCG
601 CCCGCCATAN ACGGGTTGTC TTCNCCGCG TTGCAGAACA CGACGATTTT
651 GGCGCAGCCG AAACCTTCTA GTGTGATTTC ANCCGTGCGT TTGATGGTTT
20 701 CGCCCGCCAG TCTGACCGCG TCCATATTGA TACCGCGCGC CGTACTGCCG
751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT
801 GGAACGGATN AACACCTCGT CAGAAGGCGA CATACTTTT TGCACCGCG
851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA
```

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```
1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
51  NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT
151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP*NA
30 201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP
251 ILMELHTISV VFASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
301 KVCATLT*
```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```
35      10      20      30      40      50      60
orf24a.pep  MRTAVVLLLIIMPMAASSAMMPMVCAVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
            |||||
orf24-1     MRTAVVLLLIIMPMAASSAMMPMVCAVSPGTAIISKPTQTAVMASSLSNVSTPASAAA
            |||||
40      10      20      30      40      50      60
orf24a.pep  IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP
            |||||
orf24-1     IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP
            |||||
45      70      80      90      100     110     120
orf24a.pep  TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVNVVASEFSNAAFTT
            |||||
orf24-1     TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVNVVASEFSNAAFTT
            |||||
50      130     140     150     160     170     180
orf24a.pep  PGPDTPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
            |||||
orf24-1     PGPDTPTLITASASPEPXNAPAIINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
            |||||
55      190     200     210     220     230     240
orf24a.pep  SILIPARVLPILMELHTISVVFASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
            |||||
orf24-1     SILIPARVLPILMELHTISVVFASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS
            |||||
60      250     260     270     280     290     300
```

```

      orf24a.pep      KVCATLTX
      |||||
5      orf24-1      KVCATLTX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

```

10      orf24.pep      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTQTAVMASSLSSVSTPASAAA      60
      orf24ng      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTQTAVMASSLSSVNTPASAAA      60

15      orf24.pep      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPQTLKPIXRMRATXSP      120
      orf24ng      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPQTLKPISSRMRATESP      120

      orf24.pep      TG      122
      |:
20      orf24ng      TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVRLTASEFSSAALT      180

```

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

```

      1  ATGCGCACGG  CGGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTTCGTC
      51  GGCGATGATG  CCGGAAATGG  TGTGCGCGGG  CGTGTGCGCG  GGAACGGCAA
25      101  TCATGTCCAA  ACCAACGGAG  CAGACGGCGG  TCATGGCTTC  GAGTTTGTCC
      151  AGCGTCAACA  CGCCTGCCTC  GGCGGCGGCA  ATCATACCTT  CGTCTTCGGA
      201  AACGGGGATA  AACGCGCCGC  TCAAACCGCC  GACCGCGCTG  GAAGCCATCA
      251  TGCCGCCCTT  TTTCACGGCA  TCGTTCAGCA  ATGCCAAAGC  TGCTGTTGTG
      301  CCGTGCGTAC  CGCAGACGCT  CAAGCCCATT  TCTTCAAGAA  TGCGCGCCAC
      351  CGAGTCGCCG  ACGGCGGGGG  TCGGTGCCAG  CGACAAATCG  AGAATGCCGA
30      401  ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GACCGATGAG  TTCGCCCACG
      451  CGGGTGATTT  TGAAAGCGGT  TTTCTTCACG  ACTTCGGCGA  CCTCGGTACG
      501  GCTGACCGCG  TCCGAATTTT  CCAGCGCGGC  TTTGACCACG  CCTGGACCGG
      551  ATACGCCGAC  ATTAATCACA  GCATCCGCTT  CGCCCAGAGC  GTGGAACGCA
      601  CCCGCCATAA  ACGGATTGTC  TTCCACCGCG  TTGCAGAACA  CGACGATTTT
35      651  GGCGCAGCCG  AAACCTTCGG  GTGTGATTTC  AGCCGTGCGT  TTGATGGTTT
      701  CGCCTGCCAG  CTTGACCGCA  TCCATATTGA  TACCGGCACG  CGTGTGCGG
      751  ATATTGATGG  AGCTGCACAC  GATATCGGTA  GTTTTCATCG  CTTGCGGAAC
      801  GGAACGGATC  AACACCTCAT  CCGAAGGCGA  CATACTTTT  TGCACCAGCG
      851  CGGAAAAGCC  GCCGATAAAG  GACACGCCGA  TGGCTTTGGC  TGCCTTGTCC
40      901  AAAGTCTGCG  CCACGCTGAC  ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 680>:

```

      1  MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIMSKPTE  QTAVMASSLS
      51  SVNTPASAAA  IIPSSSETGI  NAPLKPPTAL  EAIMPPFFTA  SFSNAKAAVV
45      101  PCVPQTLKPI  SSRMRATESP  TAGVGASDKS  RMPNGIFSIF  EASRPMSSPT
      151  RVILKAVFFT  TSATSVRLTA  SEFSSAALT  PGPDTPTLIT  ASASPEPWNA
      201  PAINGLSSTA  LQNTTILAQP  KPSGVISAVR  LMVSPASLTA  SILIPARVLP
      251  ILMELHTISV  VFIASGTERI  NTSSEGDIPF  CTSAEKPPIK  DTPMALAALS
      301  KVCATLT*

```

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

```

50      orf24-1.pep      10      20      30      40      50      60
      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTQTAVMASSLSSVSTPASAAA
      orf24ng      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTQTAVMASSLSSVNTPASAAA

55      orf24-1.pep      70      80      90      100     110     120
      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPQTLKPISSRMRATESP
      orf24ng      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPQTLKPISSRMRATESP
      70      80      90      100     110     120

```

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVNVVASEFSNAAFTT					
	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVRLTASEFSSAALT					
5		130	140	150	160	170	180
	orf24-1.pep	PGPDTPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
10	orf24ng	PGPDTPTLITASASPEPWNAPAINGLSSTALQNTTILAQPKPSGVISAVRLMVSPASLTA					
		190	200	210	220	230	240
	orf24-1.pep	SILIPARVLPILMELHTISVVFIA SGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
15	orf24ng	SILIPARVLPILMELHTISVVFIA SGTERTINTSSEGDIPFCTSAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
20	orf24-1.pep	KVCATLT					
	orf24ng	KVCATLT					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

```

30      1  ..ACCGACGTGC AAAAAGAGTT GGTCCGCGAA CAACGCAAGT GGGCGCAGGA
      51  AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT
      101  ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG
      151  ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG

```

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

```

35      1  ..TDVQKELVGE QRKWAQEKIS NCRQAAAQAD RQEYAEYLKL QCDTRMTRE
      51  IQYLRGYSID *

```

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

```

      1  ATGTATCGGA AACTCATTCG GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
      51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
40    101  TGCAAGGCAT ACGCGGCAAT ATTCAAGAAA CGCTCACGCA GGAAGCGCGT
      151  TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG ACAAATTAT
      201  CGCCGCCGCC TACGGTTTGG CGTTTCTTTT GGAACACGCT TCGGAAACGC
      251  AGGAAGGCGG GCGCACGTTC TGTATCGCCG ATTTGAACAT TACCGTGCCG
      301  TCTGAAACGC TTGCCGATGC CAAGGCAAAAC AGCCCCCTGT TGTACGGGGA
45    351  AACTGCTTTG TCGGATATTG TGCGGCAGAA GACGGGCGGC AATGTCGAGT
      401  TTAAAGACGG CGTATTGACG GCAGCCGTCC GCTTCCTGCC CGTCAAAGAC
      451  GGTACAGACG CATTGTGCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
      501  GTCTGCCCGC CTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
      551  GCAAGGCGGT GAAAAAGAA GACGCGGTCA GGATTTTGAG CGGAAAAGCC
50    601  CGTGAAGAAG AACCCTCCAA ACCCACGCCC GAAGACATT TGGAAACAA
      651  TGCCGCCGCG GCGGATGCGG GCGTACCCCA AGCCGAGAA GGCAGCGCCG
      701  AACCAGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
      751  GTATCACGGG CGAAGTGGA AGAGGCGCGC GTACAAAACC AGCGTGCGGA
      801  ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
55    851  AGTTGGTCCG CGAACAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
      901  CGACAAGCCG CCGCGCAGG AGACCGGCAG GAATACGCCG AATACCTCAA
      951  GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
1001  GCTATTCCAT CGATTAG

```

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQIRGN IQETLTQEAR
    51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
   101 SETLADAKAN SPILLYGETAL SDIVRQKTGG NVEFKDGLT AAVRFLPVKD
   151 GQTAFVDNTV GMAAQTLCAA LPPYGVKSIV MIDGKAVKKE DAVRILSGKA
   201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
   251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
   301 RQAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N. meningitidis*:

```

      10      20      30
   orf25.pep      TDVQKELVGEQRKWAQEKISNCRQAAAQAD
   15      orf25a      VT VSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNCRQAAAQAD
      250      260      270      280      290      300

      40      50      60
   20      orf25.pep      RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
      orf25a      RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
      310      320      330

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

   25      1 ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
      51 CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
   101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCACGCA GGAAGCGCGT
   151 TCTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
   201 CGCCGCCGCG TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
   301 AGGAAGGCGG GCGCACGTTC TGTNTCGCCG ATTTGAACAT TACCGTGCCG
   351 TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCCTGC TGTACGGGGA
   401 AACCGCTTTG TCGGATATTG TGCGGCAGAA GACGGGCGGC AATGTCGAGT
   451 TTAAGACCGG CGTATTGACG GCAGCCGTCC GCTTCCTACC CGTCAAAGAC
   501 GGTCAGANGG CATTGTGCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
   551 GTCTGCCGCG TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
   601 GCAAGGCGGT AAAAAAGAA GACGCGGTCA GGATTNTGAG CNGANAAGCC
   651 CGTGAANAAG AACCGTCCAA ANCCNNGCCC GAAGACATTT TGGAACATAA
   701 TGCCGCCGGA GGGGATGCAG ACGTACCCCA AGCCGGAGAA GACGCGCCCG
   751 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
   801 GTATCACGGG GCGAAGTGGG AGAGGCGCGN GTACAAAACC AGCGTGCGGA
   851 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
   901 AGTTGGTCCG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
   951 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
   1001 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
    51 SFAREDXXQF VDADKIIAAA XXXXXSLEHA SETQEGGRTF CXADLNITVP
   101 SETLADAKAN SPILLYGETAL SDIVRQKTGG NVEFKDGLT AAVRFLPVKD
   151 GQXAFVDNTV GMAAQTLCAA LPPYGVKSIV MIDGKAVKKE DAVRIXSXXA
   201 REXEPSKXXP EDILEHNAAG GDADVPOAGE DAPEPEILHP DDGERADTVT
   251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
   301 RQAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

   55      10      20      30      40      50      60
      orf25a.pep      MYRKLIALPFALLLAACGREPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF
      orf25-1      MYRKLIALPFALLLAACGREPPKALECANPAVLQIRGNIQETLTQEARSFAREDGRQF

```

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf25a.pep	VDADXIIAAAXXXXXSLEHASETQEGGRTFCXADLNITVPSETLADAKANSPLLYGETAL					
	orf25-1	VDADKIIAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAKANSPLLYGETAL					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf25a.pep	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQXAFVDNTVGMAAQTLAALLPYGVKSIV					
	orf25-1	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQTAFVDNTVGMAAQTLAALLPYGVKSIV					
		130	140	150	160	170	180
		190	200	210	220	230	240
15	orf25a.pep	MIDGKAVKKEDAVRIXSXXAREXEPSKXXPEDILEHNAAGGDADVPQAGEDAPEPEILHP					
	orf25-1	MIDGKAVKKEDAVRILSGKAREEPPSKPTPEDILEHNAAGGDAGVPOAAEGAPEPEILHP					
		190	200	210	220	230	240
20		250	260	270	280	290	300
	orf25a.pep	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNC					
	orf25-1	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEQRKWAQEKISNC					
25		250	260	270	280	290	300
		310	320	330	339		
	orf25a.pep	RQAAAQADRQEYAEYCLKQCDTRMTRERIQYLRGYSIDX					
	orf25-1	RQAAAQADRQEYAEYCLKQCDTRMTRERIQYLRGYSIDX					
30		310	320	330			

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

	orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAQAD	30
	orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLD	308
40	orf25.pep	RQEYAEYLKQCDTRMTRERIQYLRGYSID	60
	orf25ng	RQEYAEYLKQCDTRMTRERIQYLRGYSID	338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAGGACAT	ACGCGGCAGT	ATTcAGGAAA	CGCTcACGCA	GGAAGCGCGT
	151	TCTTTCCGCG	GCGAAGACGG	CAGGCAGTTT	TCGTATGCCG	ACAAAATTAT
50	201	CGCCGCGCTT	TACGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAACGCG
	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAc	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCCTTT	GCAGACATCG	TGCAGCAGAA	GACGGCCGCG	AATGTCGAGT
55	401	TTAAAGACGG	CGTATTGACG	CGAGCCGTCc	GCTTCTGTCc	CGCCAAAGAC
	451	GCTCGGACGG	CATTTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
	501	GTCTGCCCGC	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTTGAG	CGGCAAAGCC
60	601	CGTGAAGAAg	AACCGTCCAA	ACCCACCCCC	GAAGACATTT	TGGAACACAA
	651	TGCCGCGCGC	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCACCCG
	701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCCA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
	801	ATCCGAAGATT	ACCAAACTTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCGG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	901	cgAAGACGGC	CGCGCGACGG	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTCCAATGC	CACACGCGGA	TGACGCGCGA	ACggaTACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR
 51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
 101 SETLADAEAN SPLLYGETSL ADIVQOKTGG NVEFKDGVLT AAVRFLPAKD
 151 ARTAFIDNTV GMATQTLCAA LLPGVKSIV MIDGKAVTKE DAVRVLGSKA
 201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT
 251 VSRGEVEEAR VQNRQAESEI TKLWGGGLTD VQKELVGEQR KWAQEKISNC
 301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep	10	20	30	40	50	60
	orf25ng	10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
	orf25ng	70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
	orf25ng	130	140	150	160	170	180
25	orf25-1.pep	190	200	210	220	230	240
	orf25ng	190	200	210	220	230	240
30	orf25-1.pep	250	260	270	280	290	300
	orf25ng	250	260	270	280	290	300
35	orf25-1.pep	310	320	330	339		
	orf25ng	310	320	330			

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described
 50 above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and
 55 that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGwysGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CGsyGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTT CkGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....

10      //

     851  .....AC TTCGCTGGTA
     901  TTCCGGCGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGTCGGGCAC
     951  GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  TGTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
    151  1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
     1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
     1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
     1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
     1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CCGCGACCAC TGCTCGCCCA
    201  1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
     1351  GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCCGCGC
     1401  CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
     1451  TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
    1501  AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

30      1  MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFLVGGNPV
      51  DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSL LTYSGSN...

     251  .....TSLV
     301  FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
     351  VGEMHTGDYL STLVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
     401  IAAAMAVKVE PALII PCMSA VMAGAVCGDH CSPI SDTIL SSTGARNHI
     451  DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLFGFTTGIV LAVLIFLLKD
    501  KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

40      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGTCGGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CGGCGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTT CTGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
     301  GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGCGG CGAAAAATGCT
     351  GACCGCCTGC CTCGTGTTCC TAACCTTTAT CGACGACTAT TTCCACAGTC
     401  TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
     451  CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
     501  GCTGATGCCC GTTTCAGGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
     551  GACTGCTCGT TACCTACAAA ATCACC GAAT ACACGCCGAT GGGGACGTTT
     601  GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
     651  GTTCGTCGTC GCATGGTTTT CCTTCGACAT CGGCTCGATG GCACGTTTTCG
     701  AACAAAGCCG GTTGAACGAA GCCACGATG AAAGTGGCGT TTCAGACGCT
     751  ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC
     801  CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
     851  TCAGCATTTT GGGGGCATTG GAAAACACGG ACGTAAACAC TTCGCTGGTA
     901  TTCGGCGGCA CTTGCGGCGT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC
     951  GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  TGTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
    1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
    1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
    1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
    1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
    1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CCGCGACCAC TGCTCGCCCA
    1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```


-395-

1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
 1451 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
 1501 AAAAAACGCG CCAACGCCTG A

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
 51 DGLTHLKDMV VGLAWSGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAIA RPVTDKFKVS
 151 RTKLAYILDS TAAPMCVLMV VSSWGASIIA TLAGLLVYK ITEYTPMGTF
 10 201 VAMSLMNYA LFALIMVFV VWFSDIGSM ARFEQAALNE AHDETAVSDA
 251 TKGRVYALII PVLALIASTV SAMIYGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLA VCLTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
 351 VGEMHTGDYL STLAVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KKRANA*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

Orf26 1 MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60
 M+LID+S S +S+VP LA+ LA+ TRRV L +L V
 HI1586 14 MELIDFSSSVWSIVPALAILAIATRRLVLSLSAGIIGSLMLSDWQIGSAFNYLVKNV 73
 25 Orf26 61 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSN 97
 V L ++D + + I++F +LLG+ T+LLT SGSN
 HI1586 74 VSLVYADGEIN-SNMNIVLFLLLLGLVLTALLTVSGSN 109
 30 //
 Orf26 86 IFTSLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGAKSMFGXXXX 141
 +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
 HI1586 299 VFSVLGTFTENTVVGTSVLVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGKSMGSAIAI 358
 35 Orf26 142 XXXXXXSTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP 201
 + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP
 HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLP 418
 40 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDH CSPISDTTILSSTGARNHIDHVTSQXXXX 261
 IAAAMA P L++PC+SAVMAGAVCGDH CSP+SDTTILSSTGA+CNHIDHVT+Q
 HI1586 419 IAAAMAANAPELLLPCLSAVMAGAVCGDH CSPVSDTTILSSTGAKCNHIDHVTTQLPYA 478
 Orf26 262 XXXXXXXXXXXXXXXXXXXKSALLGFGTTGIVLAVLIFLLKDK 302
 S L GF T + L V+IF +K +
 45 HI1586 479 ATVATATSIGYIVVGFTYSGLAGFAATAVSLIVIIFAVKKR 519

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.*

meningitidis:

50 orf26.pep 10 20 30 40 50 60
 MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV
 orf26a MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
 55 10 20 30 40 50 60
 orf26.pep 70 80 90 99
 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSNXX-----
 orf26a VGLAWSGDWWSLGKPKXLVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
 60 70 80 90 100 110 120

-396-

```

orf26.pep -----
5  orf26a  LVFVTFIDDDYFHSLAVGAXARPVTDKFKVSRAKLAYILDSTAAPMCVIMPVSSWGASIIA
      130      140      150      160      170      180

10 orf26.pep -----
   orf26a  TLAGLLVTYKITEYTPMGTFVAMSLMNYYALFALIMVFVVAWFSFDIGSMARFEQAALNE
      190      200      210      220      230      240

15 orf26.pep -----
   orf26a  AHDETAVSDGSWGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
      250      260      270      280      290      300

20 orf26.pep 120      130      140      150      160      170
   orf26a  FGGTCGVFAVVLCTLGTIKTADYPKAVWQGAKSMFGAIAILILAWLISTVVGEMHTGDYL
      310      320      330      340      350      360

25 orf26a  FGGTCGVFAVVLCTLGTIKIADYPKAVWQGAKSMFGAIAILILAWLISTVVGEMHTGDYL
      310      320      330      340      350      360

30 orf26.pep 180      190      200      210      220      230
   orf26a  STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPIAAAMAVKVEPALIIPCMSA
      370      380      390      400      410      420

35 orf26a  STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPIAAAMAVKVDPSLIIPCMSA
      370      380      390      400      410      420

   orf26.pep 240      250      260      270      280      290
   orf26a  VMAGAVCGDHCSPISDTTILSSTGARCNHIDHVTSQLPYALTVAAAAASGYLALGLTKSA
      430      440      450      460      470      480

40 orf26a  VMAGAVCGDHCSPISDTTILSSTGARCNHIDHVTSQLPYALTVAAAAASGYLALGLTKSA
      430      440      450      460      470      480

   orf26.pep 300      310
   orf26a  LLGFGTTGIVLAVLIFLLKDKK
      490      500

   orf26a  LLGFGXTGIVLAVLIFLLKDKKRANAX
      490      500

```

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

```

45 1 ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
   51 TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
  101 GCATCGGTAT TCTGGTCGGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
  151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
  201 CGGCGATTGG TCGCTGGGCA AACCAAAANT CTTGGTTTTT CTGATACTTT
  251 TGGGTATTTT TACTTCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
  301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGGCG CGAAAATGCT
  351 GACCGCCTGC CTCGTGTTCG TAACCTTTAT CGACGACTAT TTCCACAGTC
  401 TCGCCGTCGG TCGCNTTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
  451 CGCGCCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCGC CTATGTGCGT
  501 GCTGATGCCC GTTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
  551 GACTGCTCGT TACCTACAAA ATCACC GAAT ACACGCCGAT GGGGACGTTT
  601 GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
  651 GTTCGTGCTC GCATGGTTCT CCTTCGACAT CGGCTCGATG GCAGGCTTCG
  701 AACAAGCCGC GTTGAACGAA GCCCACGATG AAAGTGGCGT TTCAGACGGC
  751 AGCTGGGGCA GGGTTTACGC ATTGATTATT CCCGTTTTGG CCTTAATCGC
  801 CTCAACGGTT TCCGCCATGA TCTACACCGG TGCACAGGCA AGCGAAACCT
  851 TCAGCATTTT GGGTGCAATT GAAAATACGG ACGTGAACAC TTCGTGGTGA
  901 TTCGGCGGCA CTTGCGGCGT GCTTGCCGTC GTCCTCTGCA CGCTCGGCAC
  951 GATTAAATC GCCGATTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCCA
  1001 TGTTCCGGCGC AATCGCCATT TTAATCCTTG CCTGGCTCAT CAGTAGCGTT
  1051 GTCGGCGAAA TGCACACAGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
  1101 CATCCATCCC GGCTTCCTGN CCGTCATCCT TTTCCTGCTC GCCAGCGTGA
  1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT CATGCTGCCG
  1201 ATTGCCGCGC CCATGGCGGT CAAAGTCGAT CCCTCACTGA TTATCCCGTG
  1251 TATGTCCGCC GTGATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
  1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```

-397-

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
 1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT
 1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAGAT
 1501 AAAAAACGCG CCAACGCCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
 51 DGLTHLKDMV VGLAWSGDW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RRGAKMLTAC LVEVTFIDY FHSLAVGAXA RPVTDKFKVS
 151 RAKLAYILDS TAAPMCVLMF VSSWGASIIA TLAGLLVYK ITEYTPMGTF
 10 201 VAMSLMNYA LFALIMVFV WFSFDIGSM ARFEQAALNE AHDEAVSDG
 251 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLA VCLTGTIKI ADYPKAVWQG AKSMFGAIAI LILAWLISTV
 351 VGEMHTGDYL STL VAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVD PSLIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD
 501 KKRANA*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pep	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFLVGGNPV	DGLTHLKDMV
	orf26-1	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFLVGGNPV	DGLTHLKDMV
		10	20	30	40	50	60
25	orf26a.pep	VGLAWSGDW	SLGKPKXLV	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRGAKMLTAC
	orf26-1	VGLAWSGDW	SLGKPKILV	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRGAKMLTAC
		70	80	90	100	110	120
30	orf26a.pep	LVEVTFIDY	FHSLAVGAX	ARPVTDKFK	VSRAKLAYIL	DDTAAPMCV	LMPVSSWGASIIA
	orf26-1	LVEVTFIDY	FHSLAVGAI	ARPVTDKFK	VSRTKLAYIL	DDTAAPMCV	LMPVSSWGASIIA
		130	140	150	160	170	180
35	orf26a.pep	TLAGLLVYK	ITEYTPMGTF	VAMSLMNYA	LFALIMVFV	WFSFDIGSM	ARFEQAALNE
	orf26-1	TLAGLLVYK	ITEYTPMGTF	VAMSLMNYA	LFALIMVFV	WFSFDIGSM	ARFEQAALNE
		190	200	210	220	230	240
40	orf26a.pep	AHDEAVSDG	SWGRVYALII	PVLALIASTV	SAMIYTGAQA	SETFSILGAF	ENTDVNTSLV
	orf26-1	AHDEAVSDG	SWGRVYALII	PVLALIASTV	SAMIYTGAQA	SETFSILGAF	ENTDVNTSLV
		250	260	270	280	290	300
45	orf26a.pep	FGGTCGVLA	VCLTGTIKI	ADYPKAVWQ	GAKSMFGAIA	ILILAWLIST	VVGEMHTGDYL
	orf26-1	FGGTCGVLA	VCLTGTIKT	ADYPKAVWQ	GAKSMFGAIA	ILILAWLIST	VVGEMHTGDYL
		310	320	330	340	350	360
50	orf26a.pep	STLVAGNIH	PGFLXVILF	LLASVMAFAT	GTSGWTFGIM	LPIAAAMAVK	VDPSLIIPCMSA
	orf26-1	STLVAGNIH	PGFLPVILF	LLASVMAFAT	GTSGWTFGIM	LPIAAAMAVK	VEPALIIPCMSA
		370	380	390	400	410	420
55	orf26a.pep	VMAGAVCGD	HCSPISDTT	ILSSTGARNH	IDHVTSQLPY	ALTAAAAASG	YLALGLTKSA
	orf26-1	VMAGAVCGD	HCSPISDTT	ILSSTGARNH	IDHVTSQLPY	ALTAAAAASG	YLALGLTKSA
		430	440	450	460	470	480
60	orf26a.pep	LLGFGXTGIV	LAVLIFLLK	DKKRANAX			
	orf26-1	LLGFGXTGIV	LAVLIFLLK	DKKRANAX			
		490	500				
65	orf26a.pep	LLGFGXTGIV	LAVLIFLLK	DKKRANAX			

-398-

orff26-1 |||||:|||||
 LLGFGTTGIVLAVLIFLLKDKKRANAX
 490 500

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

	orff26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFVVGGNPVDGLTHLKDMV	60
10	orff26ng	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVVGGNPVDGLTHLKDMV	60
	orff26.pep	VGLAWSDXDWSLGPVKILVFXILLGIFTSLLTYSGSN	97
15	orff26ng	VGLAWADGDWSLGPVKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
		//	
	orff26.pep	TSLVFGGTCGVFAVVLCTLGTIKTADYPKA	326
20	orff26ng	ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVLA VVLCTFGTIKTADYPKA	326
	orff26.pep	VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF	386
25	orff26ng	VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF	386
	orff26.pep	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR	446
	orff26ng	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR	446
30	orff26.pep	CNHIDHVTSQ LPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK	502
	orff26ng	CNHIDHVTSQ LPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK RADV	506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

	1	ATGCAGCTGA	TTGACTATTC	ACATTCATTT	TTCTCGGTTG	TGCCACCCTT
35	51	TTTGGCACTG	GCACTTGCCG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
	101	GCATCGGTAT	TTTGGTCCGC	GTTCGCTTTT	TGGTCCGGCG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGGCAGA
	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAAAT	CTTGGTTTTT	CTGATACTTT
40	251	TGGGCATTTT	CACTTCACTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
	301	GCCGACTGGG	CAAAACGGCA	CATTAAAAAC	CGGTGCGGCG	CGAAAATGCT
	351	GACCGCCTGC	CTCGTGTTCG	TAACCTTTAT	CGACGACTAT	TGCCACAGCC
	401	TCGCCGTCGG	TGCGATTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCAAAC	TCGCTACAT	CCTCGACTCC	ACTGCCTCGC	CCATGTGCGT
45	501	GCTGATGCCC	GTTC AAGCT	GGGCGCGGTC	GATTATCGCC	ACGCTTGCCG
	551	GATTGCTCGT	TACCTACAAA	ATTACCGAAT	ACACGCCGAT	GGGACGTTT
	601	GTCGCCATGA	GCCGTGATGA	CTATTACGCG	CTGTTTGCCC	TGATTATGGT
	651	ATTCGTGCTC	GCATGGTTCT	CCTTCGACAT	CGGCTCGATg	gCGCGTTTCG
	701	AACAGGCTGC	GTGTAACGAA	gcccaggacg	aaaccgcccgc	tTCAGACgCT
	751	ACCAAAGGTC	GTGTTTACGC	ATTGATTATT	CCCGTTTTGG	CCTTAATCGC
50	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	CGCGCAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGGGCATTT	GAAAATACCG	ACGTAAACAC	TTCCGCTGGTA
	901	TTCGGCGGCA	CTTGCGGCGT	GCTTGCCGTC	GTCTCTGCA	CGTTCGGCAC
	951	GATTAAAACC	CCCGATTATC	CCAAAAGCCGT	GTGGCAGGGT	GCGAAATCCA
55	1001	TGTTCCGGCG	AATCGCCATT	TTAATCCTCG	CCTGGCTCAT	CAGTACGGTT
	1051	GTCGGCGAAA	TGCACACGGG	CGACTACCTC	TCCACGCTGG	TTGCGGGCAA
	1101	CATCCATCCC	GGCTTCCTGC	CCGTCATCCT	CTTCCTGCTC	GCCAGCGTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCGGCAT	TATGCTGCCG
	1201	ATTGCCGCGG	CCATGGCGGT	CAAAGTCGAA	CCCGCGCTGA	TTAtcccGTG
	1251	TATGTCCGCA	GTAATGGCGG	GGGCGGTATG	CGGCGACCAC	TGTTCCGCCA
60	1301	TCTCCGACAC	GACCATCCTG	TCGTCCACCG	GCGCGCGCTG	CAACCACATC
	1351	GACCACGTTA	CCTCGCAACT	GCCTTATGCC	CTGACGGTTG	CCGCCGCCGC
	1401	CGCATCGGGC	TACCTCGCAT	TGGGTCTGAC	AAAATCCGCG	CTGTTGGGCT
	1451	TTGGCACGAC	CGGTATTGTA	TTGGCGGTGC	TGATTTTCT	GTTGAAAGAT
	1501	AAAAAACGCG	CCGACGTTTG	A		

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

-399-

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
 51 DGLTHLKDMV VGLAWADGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RCGAKMLTAC LVFVTFIDDY FHSLAVGAIA RPVTDKFKVS
 151 RAKLAYILDS TASFMCVLMF VSSWGASIIA TLAGLLVTYK ITEYTPMGTF
 5 201 VAMSLMNYA LFALIMFVV AWFSFDIGSM ARFEQAALNE AQDETAASDA
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLA VVLCFTGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
 351 VGMHTGDYL STL VAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
 401 TAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI
 10 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KKRADV*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

		10	20	30	40	50	60
15	orf26-1.pep	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFLVGGNPV	DGLTHLKDMV
	orf26ng	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFLVGGNPV	DGLTHLKDMV
		10	20	30	40	50	60
20	orf26-1.pep	VGLAWSDGDW	SLGKPKILVF	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRCGAKMLTAC
	orf26ng	VGLAWADGDW	SLGKPKILVF	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRCGAKMLTAC
		70	80	90	100	110	120
25	orf26-1.pep	LVFVTFIDDY	FHSLAVGAIA	RPVTDKFKVS	RKTALAYILD	STAAPMCVLM	PVSSWGASIIA
	orf26ng	LVFVTFIDDY	FHSLAVGAIA	RPVTDKFKVS	RKTALAYILD	STAAPMCVLM	PVSSWGASIIA
		130	140	150	160	170	180
30	orf26-1.pep	TLAGLLVTYK	ITEYTPMGTF	VAMSLMNYA	LFALIMFVV	AWFSFDIGSM	MARFEQAALNE
	orf26ng	TLAGLLVTYK	ITEYTPMGTF	VAMSLMNYA	LFALIMFVV	AWFSFDIGSM	MARFEQAALNE
		190	200	210	220	230	240
35	orf26-1.pep	AHDETAVS	DATKGRVYAL	IIPVLALIA	STVSAMIYT	GQAQSETFS	ILGAFENTDVNTSLV
	orf26ng	AQDETAAS	DATKGRVYAL	IIPVLALIA	STVSAMIYT	GQAQSETFS	ILGAFENTDVNTSLV
		250	260	270	280	290	300
40	orf26-1.pep	FGGTCGVLA	VVLCFTGTIK	TADYPKAVW	QGAQSMFGA	IAILILAWL	ISTVVGMHTGDYL
	orf26ng	FGGTCGVLA	VVLCFTGTIK	TADYPKAVW	QGAQSMFGA	IAILILAWL	ISTVVGMHTGDYL
		310	320	330	340	350	360
45	orf26-1.pep	STLVAGNIH	PGFLPVILF	LLASVMAFAT	GTSGWTFG	IMLPAAAM	AVKVEPALIIPCMSA
	orf26ng	STLVAGNIH	PGFLPVILF	LLASVMAFAT	GTSGWTFG	IMLPAAAM	AVKVEPALIIPCMSA
		370	380	390	400	410	420
50	orf26-1.pep	VMAGAVCGD	HCSPISTTIL	SSTGARNHI	DHVTSQLPY	ALTVA	AAAAASGYLALGLTKSA
	orf26ng	VMAGAVCGD	HCSPISTTIL	SSTGARNHI	DHVTSQLPY	ALTVA	AAAAASGYLALGLTKSA
		430	440	450	460	470	480
55	orf26-1.pep	LLGFGTTGIV	LAVLIFLLK	DKKRANAX			
	orf26ng	LLGFGTTGIV	LAVLIFLLK	DKKRADVX			
		490	500				
60	orf26-1.pep	LLGFGTTGIV	LAVLIFLLK	DKKRANAX			
	orf26ng	LLGFGTTGIV	LAVLIFLLK	DKKRADVX			
		490	500				
65	orf26-1.pep	LLGFGTTGIV	LAVLIFLLK	DKKRANAX			
	orf26ng	LLGFGTTGIV	LAVLIFLLK	DKKRADVX			
		490	500				

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

-400-

sp|P44263|YF86_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037
 hypothetical
 protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
 Score = 538 bits (1370), Expect = e-152
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

5
 Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXAFLVGGNPVDGLTHLKDMV 60
 M+LID+S S +S+VP LA+ LA+ TRR L +L V
 10
 Sbjct: 14 MELIDFSSSVWSIVPALLAILAIATRRVLVLSAGIIGSLMLSDWQIGSAFNYLVKNV 73

Query: 61 VGLAWADGDWSLGGPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNCRAKMLTAC 120
 V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A
 15
 Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLLGVLITALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

Query: 121 LVFVTFIDYFHS LAVGAIARPVTDKFKVSRKALAYILDSTASPMCVLMPVSSWGASIIA 180
 LVFVTFIDYFHS LAVGAIARPVTD+FKVSRKALAYILDSTA+PMCV+MPVSSWGA II
 Sbjct: 133 LVFVTFIDYFHS LAVGAIARPVTDKFKVSRKALAYILDSTAAPMCVMPVSSWGAYIIT 192

20
 Query: 181 TLAGLLVYKITEYTPMGTFVAMSLMNYALFALIMVFWVAFSFDIGSMARFEQAALNE 240
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL
 Sbjct: 193 LIGLLATYSITEYTPIGAFVAMSSMNFYAIISIIMVFFVAYFSFDIASMVRHEKLALKN 252

25
 Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA----SETFSILGAFENTDVN 296
 +D+ TKG+V LI+P+L LI +TVS MIYTGA+A + FS+LG FENT V
 Sbjct: 253 TEDQLEEETGKTGQVRNLILPILVLIITATVSMMIYTGAELAADGKVFSVLGTFENTVVG 312

30
 Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQGA KSMFGXXXXXXXXXXXXSTVVGEM 354
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M
 Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIAILFFAWTINKIVGDM 372

35
 Query: 355 HTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALI 414
 TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLPAAAMA P L+
 Sbjct: 373 QTGKYLSLTVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLPAAAMAANAPELL 432

Query: 415 IPCMSAVMAGAVCGDHCSPISD TTILSSTGARNHIDHVT SQXXXXXXXXXXXXXXXXXXXX 474
 +PC+SAVMAGAVCGDHCSP+SD TTILSSTGA+CNHIDHVT+Q
 Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSD TTILSSTGAKCNHIDHVT TQLPYAATVATATSIGYIVV 492

40
 Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501
 S L GF T + L V+IF +K +
 Sbjct: 493 GFTYSGLAGFAATAVSLVLIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,
 45 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

50
 1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
 51 CGATGAGCCT GCCAAAATTC TGA CT TGGGA TGAAAGCGGC CGATTACTCT
 101 CGGAAGTGT TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
 201 GGTCAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

55
 1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW
 51 YEDGSKKSEX VYQDDKLVRK TQWDKGYLI EP*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

60
 1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA
 101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

-401-

5
10
201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT
401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA
501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT
601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCTTGA

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>:

1 MKKLSRIVES TVLLGFSAAL PAQYYSVYFN QNGKLTATMS SAAYIROYSV
51 VAGIAHAQDF YYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
101 KMAGGFSGKG PDGEVWNWYP NGKKSAMVPY KNGLSEGTGY RYYRNGGKES
151 EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS
201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

25
30
orf27.pep
orf27a
orf27.pep
orf27a

140 150 160 170 180 190

40 50 60 70 80

200 210 220 230 240

KQWYADXS IKTEMVMVNDEPAKILTWDESG
LSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGN IKTEMVMVNDEPAKILTWDESG
RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDG YLIEPX
RLLSELSIHXXRNGVVLEWYEDGSKKXEA VYQDDKLVRKTQWDKDG YLIEPX

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

35
40
45
201 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
51 GGCCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAT CAGAACGGGA
101 AACTGACGGC GACGNTGTCT TCTGCCGNT ATATCAGGCA ATATAGTGTG
151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA
201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA
301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT
401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAACAGAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA
501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG
551 CCAAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTGTCT
601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG
651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCTTGA

This encodes a protein having amino acid sequence <SEQ ID 702>:

50
1 MKKLSRIVES TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIROYSV
51 AEGIAHAQXF XYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
101 KMAGGFSGKG PDGEVWNWYP NGKKSAMVPY KNGLSEGTGX RYYRNGGKES
151 EIQFKQNKAN GVWKQWYADG NIKTEMVMVN DEPAKILTWD ESGRLLSELS
201 IHHXNRNGVV LEWYEDGSKK XEA VYQDDKL VRKTQWDKDG YLIEP*

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

orf27a.pep MKKLSRIVFSTVLLGFSAALPAQXYSVYFNQNGKLTATXSSAAYIROYSVAEGIAHAQXF

-402-

5	orf27-1	: : MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIROYSVVAGIAHAQDF
		10 20 30 40 50 60
10	orf27a.pep	70 80 90 100 110 120 XYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGPGEWVNWYP
	orf27-1	: : YYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGPGEWVNWYP
15	orf27a.pep	130 140 150 160 170 180 NGKKSAVMPYKNGLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN
	orf27-1	: : NGKKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
20	orf27a.pep	190 200 210 220 230 240 DEPAKILTWDSEGRLLSELSIRHHXRNQGVVLEWYEDGSKKXEAQYQDDKLVRKTQWDKDG
	orf27-1	: : DEPAKILTWDSEGRLLSELSIRHHQRNNGVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG
25	orf27a.pep	YLIEPX
	orf27-1	 YLIEPX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

30 *N.gonorrhoeae*:

35	orf27.pep	KQWYADXSIKTEMVMVNDEPAKILTWDSEG	30
	orf27ng	LSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDSEG	193
	orf27.pep	RLLSELSIRHHQRNNGVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDG YLIEP	82
	orf27ng	RLLSELSIRHHKRNQGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG YLIEP	245

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

40	1	ATGAAGAAAT	TATCTCGGAT	TGTATTTTCA	ATCGTACTGT	TGGGTTTTTC
	51	GGCCGCTTTG	CCGGCGCAGA	CCTATTCTGT	TTATTTTAAAT	CAGAACGGGA
45	101	AACGTACGGC	GACGATGTCT	TCTGCCGCTT	ATATCAGGCA	ATATAGTGTG
	151	GCGGCGGGTA	TCGCACACGC	GCAGGATTTT	TATTATCCGT	CGATGAAGAA
50	201	ATATTCCGAA	CCTTATATCG	TTGCTTCAAC	GCAAATCAA	TCTTTTGTGC
	251	CTACCCGTGA	AAACGGTATG	TTGATTTTGT	GGCATTTTAA	TGGTCAGAAA
55	301	AAAATGGCGG	GGGGCTTCAG	CAAGGTAAG	CCGGACGGGG	AATGGGTCAA
	351	CTGGTATCCG	AACGGTAAAA	AATCTGCGGT	TATGCCTTAT	AAAAATGGCT
60	401	TGAGTGAGGG	TACGGGATAC	CGTTATTACC	GTAACGGCGG	CAAGGAAAGC
	451	GAAATCCAGT	TTAAGCAAAA	TAAGGCGAAC	GGCGTATGGA	AGCAATGGTA
65	501	TGCCGATGGA	AGTATCAAGA	CGGAAATGGT	TATGGTCAAC	GATGAGCCTG
	551	CCAAAATTCT	GACTTGGGAT	GAAAGCGGCC	GATTACTTTC	GGAAGTGTCT
70	601	ATCCGCCACC	ATAAACGCAA	CGGGGTGGTT	TTGGAGTGGT	ATGAAGATGG
	651	TTCTAAAAAG	AGCGAGGCTG	TTTATCAGGA	TGACAAGTTG	GTCAGGAAAA
	701	CCCAATGGGA	TAAGGATGGT	TATTTAATCG	AACCTGTA	

This encodes a protein having amino acid sequence <SEQ ID 704>:

55	1	MKKLSRIVFS	IVLLGFSAAL	PAQTYSVYFN	QNGKLTATMS	SAAYIROYSV
	51	AAGIAHAQDF	YYPSPMKKYSE	PIIVASTQIK	SFVPTLQNGM	LILWHFNGQK
60	101	KMAGGFSK GK	PDGEWVNWYP	NGKKS AVMPY	KNGLSEGTGY	RYRNRGGKES
	151	EIQFKQNKAN	GVWKQWYADG	SIKTEMVMVN	DEPAKILTWD	ESGRLLSELS
	201	IRHHKRNQGV	LEWYEDGSKK	SEAVYQDDKL	VRKTQWDKDG	YLIEP*

60 ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

	10	20	30	40	50	60
orf27-1.pep	MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIROYSVVAGIAHAQDF					

	orf27ng	MKKLSRIVFSIVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVAAGIAHAQDF	10	20	30	40	50	60
5	orf27-1.pep	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLLILWHFNQGKKMAGGFSKGPDPGEWVNWYP	70	80	90	100	110	120
	orf27ng	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLLILWHFNQGKKMAGGFSKGPDPGEWVNWYP	70	80	90	100	110	120
10	orf27-1.pep	NGKKSAMVPYKGNLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN	130	140	150	160	170	180
	orf27ng	NGKKSAMVPYKGNLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN	130	140	150	160	170	180
15	orf27-1.pep	DEPAKILTWDSEGRLLSELSIRHHQRNGVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG	190	200	210	220	230	240
20	orf27ng	DEPAKILTWDSEGRLLSELSIRHHKRNGVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG	190	200	210	220	230	240
25	orf27-1.pep	YLIEPX						
	orf27ng	YLIEPX						

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

50 This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHX LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVLGTIFWL AARIAAFIPG
101 WGASASGILG TLEFFWYGAVC MALPVIIRSON QRNYVAVFAL FVLGGTHAAF

151 HVQLHNGNLG GLLSGLQSGL VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

      1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
    51  TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101  GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151  ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201  CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGGCGGC GTTCTGGTCG
251  GCTTGACTAT CTTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
301  TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
10 351  CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAACT
401  ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGCACGCA TCGCGCGTTC
451  CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501  GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
551  TTATTTCTGT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCCAGT
15 601  CGGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGATGCCAT
651  GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCTTTG
701  CGGCAGGTGT GATTTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAACCC
751  GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
20 801  CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCCGCTTTC
851  TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901  TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CACACGGGCA ATCCGATTTA
951  TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CGGCCGTCCG TATGTTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
25 1101 GTGGAAGTAT ATTCCTTGGC TGATTCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

```

      1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFGYTGTHE LSGFYWHAHE
    51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
101  WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
151  HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIISFFTS KRLNVPQIPS
201  PKWVAQASLW LPMLTAMLMA HGVLAWSAV FAFAAGVIFT VQVYRWYKYP
251  VLKEPMLWIL FAGYLFGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
301  LGM MARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351  SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*
  
```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

```

      10      20      30      40      50      60
orf47.pep  MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYTGTTHXLSGFYWHAHEMIWGYAGLVV
      |||
orf47a     MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYTGTTHXLSGFYWHAHEMIWGYAGLVV
      |||
      70      80      90      100     110     120
orf47.pep  IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      |||
orf47a     IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      |||
      130     140     150     160     170
orf47.pep  MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGLVLM
      |||
orf47a     MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGLVLMVSGFIGLI
      |||
      130     140     150     160     170     180
orf47a     GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAFAFAAGVIFT
      |||
      190     200     210     220     230     240
  
```

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

      1 ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA
      51 TTCACTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
    101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
    151 ATGATTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
    201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGGCGGC GTTCTGGTCG
    251 GCTGACTAT CTTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
    301 TGGGGTGCCT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
    351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAATT
    401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGGCGTTC
    451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
    501 GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
    551 TTATTTCTGT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCCCACT
    601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACCGCCAT
    651 CGTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
    701 CGGCAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
    751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
    801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCCGCTTTCC
    851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTCTGACT
    901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CACACGGGCA ATCCGATTTA
    951 TCCGCCGCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
   1001 CCGCGTCCG TATGTTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
   1051 AGCATACGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
   1101 GTGGAAGTAT ATTCCTTGGC TGATTTCGTCC GCGTTCGGAC GGCAGGCCCG
   1151 GTTGA
  
```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

      1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
      51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
    101 WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
    151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GTRIISSFTS KRLNVPQIPS
    201 PKWVAQASLW LPMLTAMLMA HGVMPWLSAA FAFAGVIFT VQVYRWYKYP
    251 VLKEPMLWIL FAGYLFTGLG LIAVGASYFK PAFNLGVHL IGVGIGVLT
    301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
    351 SIRTSSVLFA LALLVYAWKY IPWLIRPSD GRPG*
  
```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```

      10      20      30      40      50      60
or47a.pep MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
or47-1 MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
      70      80      90      100     110     120
or47a.pep IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      70      80      90      100     110     120
or47-1 IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      130     140     150     160     170     180
or47a.pep MALPVIRSONQRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSLVMVSGFIGLI
      130     140     150     160     170     180
or47-1 MALPVIRSONQRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSLVMVSGFIGLI
      190     200     210     220     230     240
or47a.pep GTRIISSFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT
      190     200     210     220     230     240
or47-1 GTRIISSFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVLAWSAVFAFAAGVIFT
      250     260     270     280     290     300
or47a.pep VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
      250     260     270     280     290     300
or47-1 VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
      310     320     330     340     350     360
  
```

-406-

	orf47a.pep	LGM MARTALGHTGNPIYPPPKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA	
	orf47-1	LGM MARTALGHTGNPIYPPPKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA	
5		310 320 330 340 350 360	
	orf47a.pep	LALLVYAWKYIPWLIRPRSDGRPGX	
	orf47-1	LALLVYAWKYIPWLIRPRSDGRPGX	
10		370 380	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV	60
	ORF47ng	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV	60
20	ORF47	IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC	120
	ORF47ng	IAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC	120
25	ORF47	MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVM	172
	ORF47ng	MALPVIRSONRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVWGFGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

	1	MKFTKHPVWA	MAFRPFYSLA	ALYGALSVLL	WGFYGTGTHE	LSGFYWHAHE
	51	MIWGYAGLVV	IAFLLTAVAT	WTGQPPTRG	VLVGLTAFWL	AARIAAFIPG
30	101	WGAAASGILG	TLFFWYGAVC	MALPVIRSON	RRNYVAVFAI	FVLGGTHAA
	151	HVQLHNGNLG	GLLSGLQSG	VMVWGFGLI	GMKIIISFFTS	KRLKLPQIPS
	201	PKWVAHASLW	LPMLNAILMA	HRVMPWLSAA	FPFAAGVIFT	VQVYAGGITP
	251	IEETSCGSVA	GICYRLGNSS	G		

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

	TM segments in ORF47ng			
	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
40	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTA	CCAAACATCC	CGTCTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACGCGG	GCACTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
	101	GCTACACGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTTGGG	GTTATGCCGG	TCTCGTCGTC	ATCGCCTTCC	TGCTGACCGC
	201	CGTCGCCACT	TGGACGGGAC	AGCCGCCAC	GAGGGGCGGC	GTTCTGGTCG
50	251	GCTTGACCGC	CTTTTGGCTG	GCTGCGCGGA	TTGCCGCCTT	TATCCCGGGT
	301	TGGGGTGCGG	CGGCAAGCGG	CATACTCGGT	ACGCTGTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TtcgCAAAAC	CGGCGCAACT
	401	ATGtcgCCGT	ATTCGCAATA	TTTGTGCTGG	GCGGTACGCA	TGCGgcgTTC
	451	CACGtccAgc	tGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
55	501	GTCGGGCTTG	GTTATGGTGT	CGGGCTTTAT	CGGCCTGATT	GGGATGAGGA
	551	TTATTTTCGT	TTTTACGTCC	AAACGGTTGA	ACGTGCCGCA	GATTCCCACT
	601	CCGAAATGGG	TGGCGCAGGC	TTGCTGTGG	CTACCCATGC	TGACCGCCAT

-407-

5 651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
 701 CGGCGGGCGT GATTTTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
 751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
 801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCTGCCTTCC
 851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
 901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATTTCGATTTA
 951 TCCGCCGCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
 1001 CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
 1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
 1101 GTGGAAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCC
 1151 GTTGA

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

15 1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
 51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTAFWL AARIAAFIPG
 101 WGAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF
 151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GMRIISFFTS KRLNVPQIPS
 201 PKWVAQASLW LPMLTAILMA HGVMPWLSAA FAFAAGVIFT VQVYRWYKYP
 251 VLKEPMLWIL FAGYLFTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
 301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
 20 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

		10	20	30	40	50	60
25	orf47-1.pep	MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV 					
	orf47ng-1	MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV 					
		10	20	30	40	50	60
30	orf47-1.pep	IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC 					
	orf47ng-1	IAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC 					
		70	80	90	100	110	120
35	orf47-1.pep	MALPVIRSONRRNYVAVFAIFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI 					
	orf47ng-1	MALPVIRSONRRNYVAVFAIFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI 					
		130	140	150	160	170	180
40	orf47-1.pep	GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLAHGVLAWSAVFAFAAGVIFT 					
	orf47ng-1	GMRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVMPWLSAAFAFAAGVIFT 					
		190	200	210	220	230	240
45	orf47-1.pep	VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT 					
	orf47ng-1	VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT 					
		250	260	270	280	290	300
50	orf47-1.pep	LGMMARTALGHTGNPIYPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA 					
	orf47ng-1	LGMMARTALGHTGNPIYPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA 					
		310	320	330	340	350	360
55	orf47-1.pep	LALLVYAWKYIPWLIRPRSDGRPGX 					
	orf47ng-1	LALLVYAWKYIPWLIRPRSDGRPGX 					
		370	380				
60	orf47-1.pep	LALLVYAWKYIPWLIRPRSDGRPGX 					
	orf47ng-1	LALLVYAWKYIPWLIRPRSDGRPGX 					
		370	380				

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

65 gnl|PID|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396
 Score = 155 bits (389), Expect = 5e-37

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

```

Query: 7   PVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59
          P+W +AFRPF+  +LY L++ LW  +TG      GF      WH HEM++G+A  +
Sbjct: 14  PIWRLAFRPFFLAGSLYALLAIPLVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

Query: 60  VIAFLLTAVATWTGQPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119
          V  FLLTAV TWTGQ   G  LVGL A WLAAR+  ++ G AA   L  LF
Sbjct: 72  VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPDLLFLVALVW 130

Query: 120  CMALPVIRSONRRNYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179
          MA  +  + +RNY V  + ++ G                      +V+  + L
Sbjct: 131  MMAQMLWAVRQKRNYPIVVVLSMLGADVLIILTGLLQGNDAQRQGVLAGLWLVAALMAL 190

Query: 180  IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV----MPWLSAAFAFA 234
          IG R+I FFT + L      P W+  A L      + A+L A GV      P L  F  A
Sbjct: 191  IGGRVIPFFTQRLGKVDVAVKPVWVLDVALLVGTGVIALHAFGVAMRPQPLLGLLFV-A 249

Query: 235  AGVIFTVQVYRWYKPVLPKEPMLWILFAGYLTGLGLIAVGASYF-KPAFXXXXXXXXXXX 293
          GV  +++ RW+ K + K  +LW L  L+  +  +  +F  A
Sbjct: 250  IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWLVVAAFGLALWHFGLLAQSSPSLHALSV 309

Query: 294  XXXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353
          M+AR LGHTG + P  + AF L                      F S      +
Sbjct: 310  GSMSGLILAMIARVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAWPVGGWLW 365

Query: 354  TSSVLFALALLVYAWKYIPWLIRPSDGRPG 384
          ++V + LA  +Y W+Y P L+  R DG PG
Sbjct: 366  LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

```

35      1  ..ATGCCGCTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAAy CAGAAGyGGT
      51  AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
     101  CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT
     151  GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
     201  TGTKGCTTTC GTGATAGGSa GGTTTGyTGG kmksAsyTTG TAYrATwkkG
     40  251  CCTssCwsTG kAGmGCCkTk CkyTGGTkka swGrwArTAG TCGTGGTTTy
     301  TkTyyCACC GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
     351  CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG
     401  GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT
     45  451  TGTTTGGGTT TCTTTGTAGT TGTTGTTTAT CTCTTCAGTA ACTTTTTTAG
     501  TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT
     551  ATTCTCCAGC GCCTGAAATC ..

```

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

```

50      1  ..MPSEGS DGXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX
      51  DVETDTGDDT KTXAADXVAF VIGRFXGXXL YXXAXXXXAX XWXXXXSRGF
     101  XXHRMNLMFN VSVGDARADI GFEFIVEFEI VNGGQAERN GVEAAVSLMF
     151  CLGFFVVVVY LFSNFFSRRI TFFPFSVTGI ICRYSPAAEI ..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

-409-

	orf67.pep		MPSEGS	SDGXGXGEXEXVAHAQXDFVGFEAG	30
	orf67ng	TNFEIAVL	SGMTVRV	FYCARPAPVNGGRLKMPSEGS	146
		90	100	110	120
5	orf67.pep	VFQASPVV	TVSGVXXQLGXD	VDVETDTGDDTKTXAADXVAFVIGRFXGXXLYXXXAX	90
	orf67ng	VFQASPVV	AVAGVQGGQAGRD	VYAHARHRAEAQAAA	206
10	orf67.pep	XWXXXSRG	FXHHRMNL	MFNVSVDARADIGFEFIVEFEIVNGGQAERRNGVEAAVSLMF	150
	orf67ng	TRVGGKST	CYFFSRID	AVSDVSVGDARTDIGFEFVVEFEIVNGGQAERRNGVECAVFLMF	266
15	orf67.pep	CLGFFV	-----	VVYLFSNFFSRRITFF-PFSVTGIICRYSPAAEI	190
	orf67ng	RLLVFY	VKLVAAKS	FIIILSFQLFYVHGIFIVVFPVPTGIIRGDAPAAEVVADRH	326

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

20	1	MPSETVGSIV	NVGVD	ESVGF	SPPFPSIQHF	YRFHRIHRIR	LFRPPGPMQL
	51	NRHSHGSGNL	GRGVW	ATVLS	DKFPCGQVRI	PACAGMTNFE	IAVLSGMTVR
	101	VFYCARPAPV	NGGRL	KMPSE	GSDGIGIGES	EAVAHAQRGF	VGFEAGVFQA
	151	SPVVAVAGV	QGGQAG	RDVYA	HARHRAEAQA	AAAVAF	FLIGVFLRMSVRINR
	201	NCCVSITRVG	GKSTCY	FFSR	IDAVSDVSVG	DARTDIGFEF	VVEFEIVNGG
	251	QAERRNGVEC	AVFLM	ERLLV	FYVKLVAAKS	FIIILSFQLFY	VHGIFIVVFP
25	301	PVTGIIRGDA	PAAEV	VADRH	PGVDGMRTDV	SEIIAYRAYF	VFAWSGWFR
	351	IVGNAFGGVG	*				

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATGTTTGCTT	TTT	TAGAAGC	CTTTTTTGTC	GAATACGGTT	ATGCGGCTGT
	51	TTTTTTTGTA	TTGGT	CATCT	GCGGTTTCGG	CGTGCCGATT	CCCAGGATT
	101	TGACCTTGGT	AACAGG	CGGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
35	151	CATATTATGT	TTGCAG	TCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT
	201	CATGTTCCGC	GCCGG	ACGAA	TTTGGGGCA	GAATTCCTA	AGGTTCAAC
	251	CTATTGCGG	CATCAT	GACG	CCGTAACGTT	ATGAGCAGGT	TCAGGAAAAA
	301	TTGACAAAT	ACGGTA	ACTG	GGTCTTATTT	GTCGCCCCGT	TCCTGCCCGG
	351	TTTGAGAACG	GCCGTAT	TTG	TTACAGCCGG	TATCAGCCGC	AAGGTTTCAT
40	401	ACTTGCGTTT	TATCATT	TATG	GATGGACTGG	CCGCA...	

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFAFLEAFFV	EYGYAAVFFV	LVICGFGVPI	PEDLTLVTGG	VISGMGYTNP
	51	HIMFAVGMLG	VLVGDGIMFA	AGRIWGQXXL	XFXPIAXIMT	PXRYEQVQEK
	101	FDKYGNWVLF	VARFLPLRLT	AVFVTAGISR	KVSYLRFIIM	DGLAA...

45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATGTTTGCTT	TTT	TAGAAGC	CTTTTTTGTC	GAATACGGTT	ATGCGGCTGT
	51	TTTTTTTGTA	TTGGT	CATCT	GCGGTTTCGG	CGTGCCGATT	CCCAGGATT
	101	TGACCTTGGT	AACAGG	CGGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
	151	CATATTATGT	TTGCAG	TCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT
50	201	CATGTTCCGC	GCCGG	ACGAA	TTTGGGGCA	GAATTCCTA	AGGTTCAAAC
	251	CTATTGCGCG	CATCAT	GACG	CCGAAACGTT	ATGAGCAGGT	TCAGGAAAAA
	301	TTGACAAAT	ACGGTA	ACTG	GGTCTTATTT	GTCGCCCCGT	TCCTGCCCGG
	351	TTTGAGAACG	GCCGTAT	TTG	TTACAGCCGG	TATCAGCCGC	AAGGTTTCAT
	401	ACTTGCGTTT	TATCATT	TATG	GATGGACTGG	CCGCACTGAT	TTCCGTCCTT
55	451	ATTTGGATT	ATCTGGG	GCGA	ATACGGTGCG	CACAACATCG	ATTGGCTGAT

-410-

501 GGCAGAAATG CACAGCCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG
 551 GTGCGACCGT TGTCGCTTGG ATTTGGTGA AAAACGCCA ACGTATCCAG
 601 TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA
 651 GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
 51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP
 151 IWIYLGEYGA HNIDWLMAMK HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ
 10 201 FYRSLKEKR AQRKAATAK KAAQSKQ*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H.influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

15 Orf78: 4 FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61
 FL FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GV
 DedA: 20 FLIGFFTEYGYWAVLFLVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHMLLVSMIGV 79
 20 Orf78: 62 LVGDGIMFAAGRIWGQXXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
 L GD M+ GRI+G L F PI I+T R V+EKF +YGN VLFVARFLPGLR
 DedA: 80 LAGDSCMYWLGRIYGTKILRFPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139
 Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
 +++ +GI+R+VS+Y+RF+++D AA
 25 DedA: 140 IYMVSGITRRVS YVRFLIDFCAA 163

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

30 orf78.pep 10 20 30 40 50 60
 MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
 orf78a MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
 35 10 20 30 40 50 60
 orf78.pep 70 80 90 100 110 120
 VLVGDGIMFAAGRIWGQXXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT
 orf78a VLVGDGIMFAAGRIWGQKILFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
 40 70 80 90 100 110 120
 orf78.pep 130 140
 AVFVTAGISRKVSYLRFIIMDGLAA
 orf78a AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIA
 45 130 140 150 160 170 180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

1 ATGTTTGCCC TTTTGAAGC CTTTTTGTG GAATACGGCT ATGCGGCCGT
 51 GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
 101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
 151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
 201 CATGTTCCGC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA
 301 TTCGACAAAT ACGGCAACTG GGTGTTATTT GTCGCTCGTT TCCTGCCCGG
 351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
 401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC
 451 GTTTGGATT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT

-411-

```

501  GCGGAAAATG CACAGCCTGC AATCCGGCAT CTTTCATCGCA TTGGGCGTGC
551  TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
601  CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAACGCA AGGCGGAAAA
651  GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

```

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

```

1  MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51  HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVEVTAGISR KVSYLRFIM DGLAALISVP
151 VWIYLGEYGA HNIDWLMAMK HSLQSGIFIA LGVLAAALAW FWRKRHHYQ
10  201 LYRAQLSEKR AKRKAKEAAK KAAQKQ*

```

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

```

15  orf78a.pep      10      20      30      40      50      60
      MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
      |||:|||||
orf78-1      10      20      30      40      50      60
      MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG

20  orf78a.pep      70      80      90      100     110     120
      VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
      |||:|||||
orf78-1      70      80      90      100     110     120
      VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT

25  orf78a.pep     130     140     150     160     170     180
      AVFVTAGISRKVSYLRFIMDGLAALISVPVWVWYLYGEYGAHNIDWLMAMKHSLSQSGIFIA
      |||:|||||
orf78-1     130     140     150     160     170     180
      AVFVTAGISRKVSYLRFIIMDGLAALISVPVWVWYLYGEYGAHNIDWLMAMKHSLSQSGIFVI

30  orf78a.pep     190     200     210     220
      LGVLAAALAWFWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQX
      |||:|||||
orf78-1     190     200     210     220
      LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX
35  190     200     210     220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

```

40  orf78.pep      XXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF 137
      |||:|||||
orf78ng      YPVLVVARFLPGLRTAVFVTAGISRKVSYLRF 32

orf78.pep      IIMDGLAA 145
      :|||||
45  orf78ng      LIMDGLAALISVPVWVWYLYGEYGAHNIDWLMAMKHSLSQSGIFIALGVLAALAWFWRKR 92

```

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

```

50  1  ..YPVLVVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWYLY
51  51  GEYGAHNIDW LMAKMHSLS GIFIALGVLA AALAWFWRK RHHYQLYRAQ
101 101  LSEKRAKRKA EKAAKKAAQK Q*

```

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

```

55  1  atgtttgccc tttTggaagc CTTTTTTGTC GAAatcggCt atgcGGCCGT
51  51  GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT
101 101  TGACCTTGTT AACGGGCGGC GTGATTTTCGG GTATGGGTTA TACCAATCCG
151 151  CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT
201 201  GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC
251 251  CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA
301 301  TTCGACAAAT ACGGCAACTG GGTCTGTTT GTCGCCCGTT TCCTGCCGGG

```

-412-

5
351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
401 ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC
451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT
501 GGCAGAAATG CACAGCCTGC AATCGGGCAT CTTTCATCGCA TTGGGCGTGC
551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

10
1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51 HIMFAVGMLG VLAGDGMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP
151 VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAALAW FWRKRHHYQ
201 LYRAQLSEKR AKRKAEEKAAK KAAQKQ*

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15
orf78-1.pep 10 20 30 40 50 60
MFAFFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
|||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf78ng-1 MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
10 20 30 40 50 60
20
orf78-1.pep 70 80 90 100 110 120
VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
|||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf78ng-1 VLAGDGMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
70 80 90 100 110 120
25
orf78-1.pep 130 140 150 160 170 180
AVFVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFVI
|||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf78ng-1 AVFVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIA
130 140 150 160 170 180
30
orf78-1.pep 190 200 210 220
LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAARKAAKAAQSKQX
||: |||:||||:||||: ||:||||:||||: ||| |||||||:|
orf78ng-1 LGVLAALAWFWWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQX
190 200 210 220
35

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*:

40 sp|P45280|YG29_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA
protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
>gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212
Score = 223 bits (563), Expect = 7e-58
Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)
45 Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGVL 62
L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL
Sbjct: 21 LIGFFTEYGYWAVLFLVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGVL 80
50 Query: 63 AGDGMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122
AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +
Sbjct: 81 AGDSCMYWLGRIYGTILRFPIRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140
55 Query: 123 FVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIALG 182
++ +GI+R+VSY+RF+++D AA+ISVP+WIIYLGE GA N+DWL ++ Q I+I +G
Sbjct: 141 YMVSGITRRVSIVRFVLIDFCAAIISVPIWIIYLGEYGAHNIDWLMAKMHSLSQSGIFIALG 200
Query: 183 VL 184
L
Sbjct: 201 YL 202
60

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 87

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

1   ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAAATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA C...
```

- 15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

1   MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNAKA QTVQLEVKIA PMPAMNH...
```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

```

20   1   ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAAATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA
```

- 30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```

1   MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNAKA QTVQLEVKIA PMPAMNHGHH
151 HGEAHQH*
```

- 35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N.meningitidis*:

```

40   orf79.pep  10      20      30      40      50      60
      MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGA FMKIHNDDEAKQDFLLGGSS
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf79a   10      20      30      40      50      60
      MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGA FMKIHNDDEAKQDFLLGGSS
45   orf79.pep  70      80      90     100     110     120
      PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf79a   70      80      90     100     110     120
      PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
```

-414-

```

              130      140
orf79.pep    VTLKFKNAKAQTVQLEVKIAPMPAMNH
              |||||
5  orf79a     VTLKFKNAKAQTVQLEVKTAPMSAMDHGHGHEAHQH
              130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

      1  ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
     51  TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
10    101  AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
     151  AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCTGTTGCCG ACCGCGTCGA
    201  AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG
     251  AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
    301  AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCCA
15    351  CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCA CAAACCGTCC
     401  AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGGACCA CGGTCATCAC
     451  CACGCGAAG CGCATCAGCA CTAA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

      1  MKXLLAAVMM AGLAGAVSAA GIHVEDGWAR TVEGMKMG G AFMKIHNDEA
    20    51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
     101  SYHVMFMGXK KQLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGH
    151  HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

25    10      20      30      40      50      60
orf79a.pep MKXLLAAVMMAGLAGAVSAAGIHVEDGWAR TVEGMKMGGA FMKIHNDEAKQDFLLGGSS
      || |||||
orf79-1     MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TVEGMKIGGA FMKIHNDEAKQDFLLGGSS
      10      20      30      40      50      60

30    70      80      90     100     110     120
orf79a.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
      |||||
orf79-1     PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      70      80      90     100     110     120

35    130     140     150
orf79a.pep VTLKFKNAKAQTVQLEVKTAPMSAMDHGHGHEAHQH
      |||||
orf79-1     VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHGHEAHQH
      130     140     150
40

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45    orf79.pep    FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
      |||||
      orf79ng      INDNGVMRMREVKGVPLEAKSVTELKPGS 30

50    orf79.pep    YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
      |||||
      orf79ng      YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHGHEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

      1  ..INDNGVMRMR EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
    55    51  TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

-415-

```

1  ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTccgccgCc GGagTccAtG TCGAggACGG CTGGGCGCGC accaCTGtcg
101 aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc
151 atacaaGACt ttgtgcTCgg CGGaagcatg cccgttgccg accgcGTCGA
5  201 AGTGCAtaca cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA
251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCACG TGATGTTTAT GGGTTTGAAG AAACAAGTGA AAGAGGGCGA
151 351 CAAGATTCCC GTTACCTGA AATTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAAACCGCG CCGATGTCCG CAATGAACCA CGGTTCATCAC
10 451 CACGGCGAAG CGCATCAGCA CTAA

```

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

```

1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TVEGMKMG GAFMKIHND EAFM
51 IQDFVLGGSM PVADRVEVHT HINDNGVM RMREVKG GVPLEAKS VTELKPG
151 101 SYHVMFMGLK QLKEGDKIP VTLKFKNA KAQTQVQ LEVKTA PMSAMNH GHH
151 HGEAHQH*

```

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

```

10 20 30 40 50 60
orf79-1.pep MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TVEGMKIGGAFMKIHND EAKQDFLLGGSS
20 orf79ng-1 MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TVEGMKMGGA FMKIHND EAIQDFVLGGSM
10 20 30 40 50 60
orf79-1.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLK KQLKEGDKIP
25 orf79ng-1 PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLK KQLKEGDKIP
70 80 90 100 110 120
orf79-1.pep VTLKFKNAKAQTQVQLEVKIAPMPAMNHGHHHGEAHQH X
30 orf79ng-1 VTLKFKNAKAQTQVQLEVKIAPMPAMNHGHHHGEAHQH X
130 140 150

```

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

```

35 gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
   Score = 63.6 bits (152), Expect = 6e-10
   Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

40 Query: 24 VEDGWAR TVEGMKMGGA FMKIHND EAIQDFVLGGSM PVADRVEVHTHINDNGVMRMREV 83
   V+ W G M I N+ D+++G +A RVE+H + +N V +M
   Sbjct: 27 VKHPWVMEPPPGPNTTMMGMII VNEGDEPDYLIGAKTDIAQRVELHKT VTIENDVAKMVPQ 86

   Query: 84 KGGVPLEAKSVTELKPGSYHVMFMGLK KQLKEGDKIPVTLKFKNAKAQTQVQLEV 137
   + + + K E K YHVM +GLKK++KEGDK+ V L F+ + TV+ V
45 Sbjct: 87 ER-IEIPPKGKVEFKHHGYHVM IIGLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```

5      1  ATGACGGTAA CTGCGGCCGA AGGCGGCCAA GCTGCCAAGG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGCGGCCGCA ATATGTTTGT GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
    10  301  CGGATTCCGG TTGTGAAATC CATCTATTCT AGTGTGAAAA AAGTATCCGA
     351  ATAcgTGCTG TCCGACAGCA GCCGTTCTGT TAAACGCCG GTACTCGTGC
     401  CGTTTCCCCA GCCCGGTATT TGGACGATyG CTTTCGTGTC AGGGCAGGTG
     451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTTCCGT
     501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
    15  551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AsCATTGAAA
     601  TATGTGATT CGCTGGGTAT GGTcATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
     701  AA

```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```

20      1  MTVTAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL
      51  PKQWRPQYVL GFNIPGLGVI VAIaVLFVTG LFAANVLGRQ ILAAWDSLLG
     101  RIPVVKSIYS SVKKVSEYVL SDSSRSFKTP VLVFPQPQGI WTIAFVSGQV
     151  SNAVKAALPX DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEXLK
     201  YVISLGMVIP DDLPVKTLAX PMPSEKADLP EQQ*

```

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

```

30      1  ATGACGGAAC nTGCGGCCGA AGGCGGCCAA GCTGCCAArG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGCGGCCGCA ATATGTTTGT GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
     301  CGGATTCCGG TTGTGAAATC CATCTATTCT AGTGTGAAAA AAGTATCCGA
     351  ATCGCTGCTG TCCGACAGCA GCCGTTCTGT TAAACGCCG GTACTCGTGC
     401  CGTTTCCCCA GCCCGGTATT TGGACGATTG CTTTCGTGTC AGGGCAGGTG
     451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAG GACGGCGATT ATCTTTCCGT
    35  501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
     551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCATTGAAA
     601  TATGTGATT CGCTGGGTAT GGTcATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
    40  701  AA

```

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```

45      1  MTEXAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL
      51  PKQWRPQYVL GFNIPGLGVI VAIaVLFVTG LFAANVLGRQ ILAAWDSLLG
     101  RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPQPQGI WTIAFVSGQV
     151  SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
     201  YVISLGMVIP DDLPVKTLAG PMPSEKADLP EQQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N. meningitidis*:

```

50      10      20      30      40      50      60
    orf98.pep  MTVTAAEGGKA AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL PKQWRPQYVL
              || |||||
    orf98a     MTEPAAEGGKA AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL PKQWRPQYVL
              10      20      30      40      50      60

```

-417-

```

      70      80      90      100      110      120
orf98.pep  GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL
5 orf98a    GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSXSLL
      70      80      90      100      110      120

      130      140      150      160      170      180
orf98.pep  SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNAVKAALPKDGDYLSVYVPTTPNPTGGYY
10 orf98a    SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNAVKAALPKDGDYLSVYVPTTPNPTGGYY
      130      140      150      160      170      180

      190      200      210      220      230
orf98.pep  IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAXPMPSEKADLPEQQX
15 orf98a    IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAGPMPSEKADLPEQQX
      190      200      210      220      230

```

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

```

20      1  ATGACGGAAC  CTGCGGCCGA  AGGCGGCAAA  GCTGCCAAGG  CGTTAAAAAA
51  ATATCTGATT  ACGGGCATTT  TGGTCTGGCT  GCCGATTGCG  GTAACGGTTT
101 GGGTGGTTTC  CTATATCGTT  TCCGCGTCCG  ATCAGCTCGT  CAACCTGCTG
151 CCGAAGCAAT  GCGGCCGCA  ATATGTTTTG  GGGTTAATA  TCCCGGGGCT
201 GGGCGTTATC  GTTGCCATTG  CCGTATTGTT  TGTAAACCGG  TTATTGCGCG
25 251 CAAACGTATT  GGGCCGGCAG  ATTCTTGCCG  CGTGGGACAG  CTTGTTGGGG
301 CGGATTCCGG  TTGTGAAGTC  CATCTATTCG  AGTGTGAAAA  AAGTATCCGA
351 NTCGTTGCTG  TCCGACAGCA  GCCGTTCTGT  TAAAACACCA  GTACTCGTGC
401 CGTTTCCCCA  ATCGGGTATT  TGGACAATCG  CATTCGTGTC  CGGTCAAGTG
451 TCGAATGCGG  TTAAGGCCGC  ATTGCCGAAG  GACGGCGATT  ATCTTCCGT
30 501 GTATGTTCCG  ACCACGCCGA  ATCCGACCGG  CGGTACTAT  ATTATGGTAA
551 AGAAAAGCGA  TGTGCGCGAA  CTCGATATGA  GCGTGGACGA  AGCGTTGAAA
601 TATGTGATT  CGCTGGGTAT  GGTCAATCCCT  GACGACCTGC  CCGTCAAAAC
651 ATTGGCAGGA  CCTATGCCGT  CTGAAAAGGC  GGATTGCCC  GAACAACAAT
701 AA

```

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

```

      1  MTEPAAEGGK  AAKALKKYLI  TGILVWLPIA  VTVWVVSIV  SASDQLVNLL
51  PKQWRPQYVL  GFNIPGLGVI  VAIAVLFVTG  LFAANVLGRQ  ILAAWDSLLG
101 RIPVVKSIYS  SVKKVSXSL  SDSSRSFKTP  VLVPFPQSGI  WTIAFVSGQV
151 SNAVKAALPK  DGDYLSVYVP  TTPNPTGGYY  IMVKKSDVRE  LDMSVDEALK
40 201 YVISLGMVIP  DDLPVKTLAG  PMPSEKADLP  EQQ*

```

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

```

      10      20      30      40      50      60
orf98a.pep  MTEPAAEGGKAAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL
45 orf98-1    MTEXAAEGGKAAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL
      10      20      30      40      50      60

      70      80      90      100      110      120
orf98a.pep  GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSXSLL
50 orf98-1    GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESLL
      70      80      90      100      110      120

      130      140      150      160      170      180
orf98a.pep  SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNAVKAALPKDGDYLSVYVPTTPNPTGGYY
55 orf98-1    SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNAVKAALPKDGDYLSVYVPTTPNPTGGYY
      130      140      150      160      170      180

      190      200      210      220      230
orf98a.pep  IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAGPMPSEKADLPEQQX
60 orf98-1    IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAGPMPSEKADLPEQQX
      190      200      210      220      230

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

5	orf98.pep	MTVTAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
10	orf98.pep	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL	120
	orf98ng	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLLXRIIPVVKSIYSSVKKVSESL	120
15	orf98.pep	SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNVKAALPXDGDYLSVYVPTTNPNTGGYY	180
	orf98ng	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPDGDYLSVYVPTTNPNTGGYY	180
	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAXPMPSEKADLPEQQ	233
	orf98ng	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELPEQQ	233

- 20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

1	MTEPAAEGGK	AAKALKKYL	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLX
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPFPQSGI	WTIAFVSGQV
151	SNAVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DDLVPKTLAG	PMPPEKAELP	EQQ*	

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTTGT	CAACCTGCTG
151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTGT	GGGTTTAATA	TCCCGGGCT
201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTGGCCG
251	CAAACGTGTT	GGCGGCCGAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTgggg
301	cggattcccg	TTGTCAAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
351	ATCGCTGCTG	TCCGACAGCA	GCCGTTTCGT	TAAACGCCG	GTAAGTCTGC
401	CGTTTCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
451	TCCAATGCGG	TTAAGGCCGC	ATTGCCGCG	GATGGCGATT	ATCTTTCGGT
501	GTATGTCCCG	ACCACGCCCA	ACCCGACCGG	CGGTTACTAT	ATTATGGTAA
551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
601	TATGTGATT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCCC	GAACAACAAT
701	AA				

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

1	MTEPAAEGGK	AAKALKKYL	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPFPQSGI	WTIAFVSGQV
151	SNAVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DDLVPKTLAG	PMPPEKAELP	EQQ*	

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

50	orf98-1.pep	MTExAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng-1	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
55	orf98-1.pep	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	120

-419-

orf98ng-1		GFNIPGLGVIVAIAVLVFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	70	80	90	100	110	120
			130	140	150	160	170	180
5	orf98-1.pep	SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTTPNPTGGYY						
	orf98ng-1	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPQDGDYLSVYVPTTTPNPTGGYY	130	140	150	160	170	180
			190	200	210	220	230	
10	orf98-1.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPSEKADLPEQQX						
	orf98ng-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELEPEQQX	190	200	210	220	230	

- 15 Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 89

- 20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

	1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTA CT CGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	G ₅ GgTACTCA
25	201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGChCGTAA	AGGCCkCAAG
	251	ssCGsGCTTG	CCTTGAACAA	GGCGGGTTTG	GCGTATTTTG	AAGGGCGTTT
	301	TGAAAAGGCG	GAAGTAGAAG	CCTCACGCGT	GTTGGTCAAC	AAAGtAGGCC
	351	GaGAGACAAC	CGGACTTTGG	CATTGATGCT	GrGCGCGCAC	GCCGCCGAC
	401	AGATGGAAAA	CATCGAssTG	CGCGACCGTT	ATCTTGCGGA	AATCGCCAAA
30	451	CTGCCGGA	AACAGCAGCT	TTCCCGTTAT	CTTTTGTGG	CGGAATCGGC
	501	GTTGAACCGG	CGCGATTACG	AAGCGGCGGA	AGCCAATCTT	CATGCGGCGG
	551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTCTTAC
	601	GCTTTCGACA	GGGGCGACGC	GTTGCAGGTT	CTGGCAAAAA	CCGAAAAACT
	651	TTCCAAGGCG	GGCGCGTTGG	GCAAATCGGA	AATGGAACGG	TATCAAAATT
35	701	GGGCATATCC	GTGCGCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
	751	AACCTGCCTG	AAGCGGATTTC	CCGACAGCCT	CAAAAACGGG	GAATTGAGCG
	801	TATCGGTTGC	GGAAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGTC
	851	AAATGGGTCA	AACAGCATT	TCCGCAsAAC	CGCCGCCCG	AGCTTTTGGA
	901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
40	951	CCATCGATTT	TGCGATGCT	TGGCTGAAAG	AACAGCCCGA	TAACGCGCTT
	1001	CTGCTGATGT	ATCTCGGTCG	GCTCGCCTTC	GGCCGCAAC	TTTGGGGCAA
	1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTGC	ATTAAAGCCG	AGTATTTCCG
	1101	CGCGTTTGGT	TCTAACAAAG	GTTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
	1151	GCGGAGGCGC	AC...			

- 45 This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

	1	MKTVVWIVVL	FAAAVGLALA	SGIYTGdVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNIPE	KMQRFGSARK	GXKXXLALNK	AGLAYFEGRF
	101	EKAELASRV	LVNKVGRDNR	TLALMLXAHA	AGQMenIXXR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLXIRYA
50	201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	DAADAALKT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	XNRRPELLEA
	301	FVESVRFLGE	REQQKAIDFA	DAWLKEQPDN	ALLMYLGRLL	AFGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	TKVFDEIGEP	QKAEAH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

55	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTA CT CGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GCTACTCA

-420-

5
10
15
20

```

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG
251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTTT
301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG
351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA
401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
451 CCGGAAAAC AGCAGCTTTC CCGTTATCTT TTGTGGCGG AATCGGCGTT
501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAATTCTT
651 CAAGCGGGG GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
701 CATACCGCCG CCAGCTGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACGGGGAAT TGAGCGTATC
801 GGTTCGCGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
901 TTTGTCGAAA GCGTGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
951 CGATTTGCC GATGCTTGGC TGAAGAACA GCCCGATAAC GCCTTCTGTC
1001 TGATGTATCT CGGTCGGCTC GCCTACGGCC GCAAACCTTG GGGCAAGGCA
1051 AAAGGCTACC TTGAAGCGAG CATTCGATTA AAGCCGAGTA TTTCCGCGCG
1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACCG CAGAAGGCGG
1151 AGGCGCAGCG CAACTTGTT TTGGAAGCCG TCTCCGATGA CGAACGTAC
1201 GCAGCGTTAG AGCAGCATAG CTGA

```

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

25
30

```

1 MKTVVWIVVL FAAAVGLALA SGIYTG DVYI VLGQTMLRIN LHAFVLGSLI
51 AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNANLT RLVLRLQRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMER YQNWAYRRQLA DAADAAALKT
251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH HNRRLPELLEA
301 FVESVRFGE REQQAIDFA DAWLKEQPDN ALLMYLGR LAYGRKLWGKA
351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRLNV LEAVSDDERH
401 AALEQHS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

35 *meningitidis*:

40
45
50
55
60

```

          10      20      30      40      50      60
orf100.pep MKTVVWIVVLF FAAAVGLALASGIYTG DVYI VLGQTMLRIN LHAFVLGSLI AVVVWYFLFK
          10      20      30      40      50      60
orf100a     MKTVVWIVVLF FAAAXGLALASGIYTG DVYI VLGQTMLRIN LHAFVLGSLI AVVVWYFLFK

          70      80      90      100     110     120
orf100.pep FIIGVLNIPEKMQRFGSARKGXKXXLALNKAGLAYFEGRFEKAELEASRVLVNKKVGRDNR
          70      80      90      100     110     120
orf100a     FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR

          130     140     150     160     170     180
orf100.pep TLALMLXAHAAAGQMENIXRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH
          130     140     150     160     170     180
orf100a     TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH

          190     200     210     220     230     240
orf100.pep AAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMER YQNWAYRRQLA
          190     200     210     220     230     240
orf100a     AAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMER YQNWAYRRQLX

          250     260     270     280     290     300
orf100.pep DAADAAALKTCLKRIPDSLKNGELSVSVAE KYERLGLYADAVKWKQHYH PKNRRPELLEA
          250     260     270     280     290     300
orf100a     DAADAAALKTCLKRIPDSLKNGELSVSVAE KYERLGLYADAVKWKQHYH PHNRRPELLEA

```

-421-

```

      310      320      330      340      350      360
orf100.pep FVESVREFLGEREQQKAIDFADAWLKEQPDNALLMYLGR LAFGRKLWGKAKGYLEASIAL
5 orf100a FVESVREFLGERDQQKAIDFADAWLKEQPDNALLXYLGR LAYGRKLWGKAKGYLEASIAL
      310      320      330      340      350      360

      370      380
orf100.pep KPSISARLVLTkVFDEIGEPQKAEAH
10 orf100a KPSISARLVLAkVFDETGEPOKAEQRNLVLASVAEENRPSAETHX
      370      380      390      400

```

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

```

15 1 ATGAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CNNTCGGGCT
51 GGCATTGGCG TCGGGCATTN ACACCGGCGA CGTGTATATC GTACTCGGAC
101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTAGG TTCGCTGATT
151 GCCGTCTGGT TGTGGTATTT CCTGTTCAA TTTATCATCG GCGTACTCAA
201 TANCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG
251 CCGCGCTTGC TTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT
20 301 GAAAAGGCGG AACTTGAAGC CTCGCGCGTA TTGGGAAACA AAGAGGCGGG
351 GGATAACCGG ACTTTGGCAT TGATGTTGGG CGCACATGCC GCCGGGCGAG
401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
451 CCGGAAAAGC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT
501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
25 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAANTTTC
651 CAAGGCGGGC GCGTNGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
701 CATACCGCCG CCAGCTGNCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCTGAAGC GGATTCCCGA CAGCCTCAA AACGGGAAT TGAGCGTATC
30 801 GGTGCGGAA AAGTACGAAC GTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GACCCGAAT TTTGGAAGCN
901 TTTGTCGAAA GCGTGCGCTT TTTGGGCGAA CGCGATCAGC AGAAAGCCAT
951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAT GCGCTTCTGC
35 1001 TGANGTATCT CGTGCGGCTC GCCTACGGCC GCAAACTTTG GGGCAAGGCA
1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG
1101 TTTGGTTCTG GCAAAGGTTT TTGACGAAAC CGGAGAACCG CAGAAGGCGG
1151 AGGCGCAGCG CAACTTGGTT TTGGCAAGCG TTGCCGAGGA AAACCGNCCT
1201 TCCGCCGAAA CCCATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 754>:

```

40 1 MKTVVWIVVL FAAAXGLALA SGIXTGDVYI VLGQTMLRIN LHAFVLGSLI
51 AVVVWYFLFK FIIGVLNXPE KMQRFGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAMNANLT RLVRLLQRYA
45 201 FDRGDALQVL AKTEKXSKAG AXGKSEMERY QNWAYRRQLX DAADAAALKT
251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNRPELLEA
301 FVESVREFLGE RDQQAIDFA DAWLKEQPDN ALLLXYLGR LAYGRKLWGKA
351 KGYLEASIAL KPSISARLV LAKVFDETGE POKAEQRNLV LASVAEENRP
401 SAETH*

```

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

```

50 orf100a.pep      10      20      30      40      50      60
      MKTVVWIVVLF AAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
      orf100-1      MKTVVWIVVLF AAAXGLALASGIYTG DVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
      10      20      30      40      50      60

55 orf100a.pep      70      80      90      100     110     120
      FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRF EKAELEASRV LGNKEAGDNR
      orf100-1      FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRF EKAELEASRV LGNKEAGDNR
      70      80      90      100     110     120

60 orf100a.pep      130     140     150     160     170     180
      TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAAEANLH
      orf100-1      TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAAEANLH
      130     140     150     160     170     180

```

-422-

		130	140	150	160	170	180
5	orf100a.pep	190	200	210	220	230	240
	orf100-1	190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVLF	AAAVGLALASGIY	TGDVYIVLGQT	MLRLNLHAFV	LGSLIAVVVW	YFLFK	60
	orf100ng	MKTVVWIVVLF	AAAVGLALASGIY	TGDVYIVLGQT	MLRLNLHAFV	LGSLIAVVVW	YFLFK	60
35	orf100.pep	FIIGVLNIPEK	MQRFGSARKG	XKXXLALNKAG	LAYFEGRFEK	AEEASRVLVN	KVGRDNR	120
	orf100ng	FIIGVLNIPEN	MRRSGSARKG	RKAALALNKAG	LAYFEGRFEK	AEEASRVLVN	KEAGDNR	120
40	orf100.pep	TLALMLXAHAA	QOMENIXXRDR	YLAEIAKLPEK	QQLSRYLLLA	ESALNRRDYE	AAEANLH	180
	orf100ng	TLALMLGAHAA	QOMENIELRDR	YLAEIAKLPEK	QQLSRYLLLA	ESALNRRDYE	AAEANLH	180
45	orf100.pep	AAAKMNANLTRL	VRLXIRYAFDR	GDALQVLAKTE	KLKAGALGKSE	MERYQNWAYR	RQLA	240
	orf100ng	AAAKMNANLTRL	VRLQRLRYAFD	RGDALQVLAKTE	KLKAGALGKSE	MERYQNWAYR	RQMA	240
50	orf100.pep	DAADAAALKTCL	KRIPDSLKNGEL	SVSVAEKYERL	GLYADAVKWVK	QHYPXNRRPE	LLA	300
	orf100ng	DAADAAALKTCL	KRIPDSLKNGEL	SVSVAEKYERL	GLYADAVKWVK	QHYPHNRRPE	LLA	300
55	orf100.pep	FVESVRFLGERE	QQKAIDFADAW	LKEQPDNALLM	YLGRLA	FGRKLWGKAK	GYLEASIAL	360
	orf100ng	FVESVRFLGERE	QQKAIDFADSW	LKEQPDNALLM	YLGRLAYGRK	LWGKAKGYLE	ASIAL	360
	orf100.pep	KPSISARLVLT	TKVFDEIGEPQ	KAEAH				386
	orf100ng	KPSIPARLVLA	KVFDETAQSQK	AEQARNLVLAS	VAGENRPSA	ETR		405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

	1	ATGAAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT	
60	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCCG	CGTGTATATC	GTA	CTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TT	CGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTTAAA	TTCATCATCG	GCG	TACTCAA
	201	TATCCCCGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCC	GCAAGG
	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCTG	AGGG	CGTTTT
	301	GAAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAG	CCCGG
65	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGG	ACAGA
	401	TGAAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCC	AACTG

-423-

5
10
15

```

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGCGTT
501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
601 TTCGATCGGG GCGATGCGTT GCAGGTCTTG GCAAAAaccG AAAAACTTTC
651 CAAGCGGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
701 CATACCGCCG CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACGGGGAAT TGagcGTATC
801 GGTTCGGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
901 TTTGTCGAAA GCGTGCGCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
951 CGATTTTGCC GATTCTTGGC TGAAGAACA GCCCGATAAC GCGCTTCTGC
1001 TGATGTATCT CGGCCGGCTC GCCTACGCC GCAAACCTTG GGGTAAGGCA
1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCCGCGCGC
1101 TTTGGTGTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG
1151 AAGCACAGCG CAACTTGGT TTGGCAAGCG TTGCCGGGA AAACCGCCCT
1201 TCCGCCGAAA CCCGTGA

```

This encodes a protein having amino acid sequence <SEQ ID 756>:

20
25

```

1  MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLgQTMlRIN LhAFVLGSli
51  AVVVWYFLFK FIIGVLNIPE NMRRSGSARK GRKAALAlNK AGLAYFEGRF
101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMenIELR DRYLAElAKL
151 PEKQQLSRYL LLAESAlNRR DYEAeANLH AAaKMnANL RLVRlQLRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAAALKT
251 CLKRIPDSLK NGELSVsVAE KYERlGLYAD AVKwVKQHYP HNRrPELLEA
301 FVESVRFLGE REQQKAIDFA DSWLKEQPDN ALLlMYLGRL AYGRKLWGKA
351 KGYLEASIAL KPSIPARLV AKVFDETAQS QKAEaQRNLV LASVAGENRP
401 SAETR*

```

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

30
35
40
45
50
55
60
65

```

          10      20      30      40      50      60
orf100-1.pep  MKTVVWIVVLFAAAVGLALASGIYTGdVYIVLgQTMlRINLhAFVLGSliAVVVWYFLFK
          |||
orf100ng      MKTVVWIVVLFAAAVGLALASGIYTGdVYIVLgQTMlRINLhAFVLGSliAVVVWYFLFK
          10      20      30      40      50      60

          70      80      90      100     110     120
orf100-1.pep  FIIGVLNIPEKMQRFGSARKGRKAALAlNKAGLAYFEGRFEKAELEASRVLGnKEAGDNR
          |||
orf100ng      FIIGVLNIPENMRRSGSARKGRKAALAlNKAGLAYFEGRFEKAELEASRVLGnKEAGDNR
          70      80      90      100     110     120

          130     140     150     160     170     180
orf100-1.pep  TLALMLGAHAAGQMenIELRDRYLAElAKLPEKQQLSRYLlLAESAlNRRDYEAeANLH
          |||
orf100ng      TLALMLGAHAAGQMenIELRDRYLAElAKLPEKQQLSRYLlLAESAlNRRDYEAeANLH
          130     140     150     160     170     180

          190     200     210     220     230     240
orf100-1.pep  AAaKMnANLRLVRlQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA
          |||
orf100ng      AAaKMnANLRLVRlQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA
          190     200     210     220     230     240

          250     260     270     280     290     300
orf100-1.pep  DAADAAALKTCLKRIPDSLKNGELSVsVAEKYERlGLYADAVKwVKQHYPHNRrPELLEA
          |||
orf100ng      DAADAAALKTCLKRIPDSLKNGELSVsVAEKYERlGLYADAVKwVKQHYPHNRrPELLEA
          250     260     270     280     290     300

          310     320     330     340     350     360
orf100-1.pep  FVESVRFLGEREQQAIDFADAWLKEQPDNALLlMYLGRLAYGRKLWGKAkGYLEASIAL
          |||
orf100ng      FVESVRFLGEREQQAIDFADAWLKEQPDNALLlMYLGRLAYGRKLWGKAkGYLEASIAL
          310     320     330     340     350     360

          370     380     390     400
orf100-1.pep  KPSISARLVlAKVFDEIGEPQKAeAQRNLVLEAVSDDERHAAlEQHSX
          |||
orf100ng      KPSIPARLVlAKVFDETAQSQKAeAQRNLVlLASVAGENRPsaETR*

```

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTTCGTG
      51  GTTTGCAGGG CTGTTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
     201  CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
     251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
     301  GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATsTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20      1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTTCGTG
      51  GTTTGCAGGG CTGTTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
     201  CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
     251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
     301  GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35      1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40      orf102  3  FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      HP1484  8  FLWVKAFHVIAVISWMAALFYLPRL FVYHAENAHKKEFVGVVQIQEK--KLYSFIASPAM 65
      orf102  63  GAVVFGAAIPFAAG--WWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY 119
      HP1484  66  GFTLITGILMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRNARFY 125
      orf102  120 RVFNEIPXXXXXXXXXXXXXFKPF 142
      HP1484  126 RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N. meningitidis*:

5		10	20	30	40	50	60
	orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYR	FMSPL		
	orf102a	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYR	FMSPL		
10		70	80	90	100	110	120
	orf102.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
	orf102a	GFGAVVFGAAIPFAAGWWSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
15		130	140				
	orf102.pep	VFNEIPVLLMVAALYXVVF	KPF				
20		130	140				
	orf102a	VFNEIPVLLMVAALYL	VVF				

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

	1	ATGATGTTT	CTTGGTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTTTCGTG
	51	GTTTGCAGGG	CTGTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
25	101	TTGATGTGCC	CGCGGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
	151	GTGCGGCTGT	ACCGTTTAT	GTCGCCGTTG	GGCTTCGGCG	CGGTCTGTGT
	201	CGGCGCGGCG	ATACCGTTT	GTCGCCGCTG	GTGGGGCAGC	GGCTGGGTAC
	251	ACGTCAAAC	GTGTTGGGC	TTGATGCTCT	TGGCTTACCA	GTTGTATTGC
	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
30	351	CTGGTACCGC	GTGTTCAACG	AAATCCCCGT	GCTGCTGATG	GTTGCCGCGC
	401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 762>:

	1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVNMAMIDVPRGN	PEYVRLSGMA
	51	VRLYRFMSPL	GFGAVVFGAA	IPFAAGWWS	GWVHVKLCGL	MLLAYQLYC
35	101	GVLLRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYL	VVF

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYR	FMSPL		
40	orf102-1	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYR	FMSPL		
		70	80	90	100	110	120
	orf102a.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
45	orf102-1	GFGAVVFGAAIPFAAGWWSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
		130	140				
50	orf102a.pep	VFNEIPVLLMVAALYL	VVF				
	orf102-1	VFNEIPVLLMVAALYL	VVF				

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:

-426-

```

    orf102.pep  MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL  60
    orf102ng    MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL  60

5    orf102.pep  GFGAVVFGAAIPFAAGWGWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120
    orf102ng    GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120

10   orf102.pep  VFNEIPVLLMVAALYXVVFKPF  142
    orf102ng    VFNEIPVLLMVAALYLVVFKPF  142
  
```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1    ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGT
51   GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
15  101  TTGATGCGCC GCGCGGCAAT CCCGAGTATG TGCGCCTGTC GGGGATGGCG
    151  GTGCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTCGCGC CGGTCGTGTT
    201  CGGCGCGGCG ATACCGTTTG CCGCcggcg GTGGGGCagc ggctggGTTT
    251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTATCA GTTGATTGTC
    301  GGCCTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
20  351  CTGGTACCGC GTGTTCAACg aAATCCCCGT GCTGCTGATG GTTGCCGCGC
    401  TGTATCTGGT CGTGTTCAAA CCGTTTTGA
  
```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1    MMFSWFKLFH LFFVISWFAG LFYLPRIFFV MAMIDAPRGN PEYVRLSGMA
25  51   VRLYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHVKLCLG LMLLAYQLYC
    101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVEFK PF*
  
```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

                                10      20      30      40      50      60
    orf102-1.pep  MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
30   orf102ng    MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
                                10      20      30      40      50      60

                                70      80      90      100     110     120
    orf102-1.pep  GFGAVVFGAAIPFAAGWGWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
35   orf102ng    GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
                                70      80      90      100     110     120

                                130     140
40   orf102-1.pep  VFNEIPVLLMVAALYLVEFKPF
    orf102ng      VFNEIPVLLMVAALYLVEFKPF
                                130     140
  
```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45   gi|2314656 (AE000647) conserved hypothetical integral membrane protein
    [Helicobacter pylori] Length = 148
    Score = 79.2 bits (192), Expect = 1e-14
    Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

50   Query: 3    FSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
    F W K FH+  VISW A LFYLPRIFFV A      +      V++      +LY F++
    Sbjct: 8    FLWVKAFHVIIVISWMAALFYLPRLFYVHAENAHKKEFVGGVVQIQEK--KLYSFIASPA 65

55   Query: 63   GAVVFGAAIP-----FAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFS 115
    G +      +      F +G      GW+H KL L ++LLAY YC  +R  +      +
    Sbjct: 66   GFTLITGILMLLIEPTLFKSG----GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121

    Query: 116  HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
    R+YRVFNE P      KPF
60   Sbjct: 122 ARFYRVFNEAPTILMILIVILVVVKPF 148
  
```


Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1  ATGGCAAAAA  TGATGAAATG  GCGCGCTGTT  GCGGCGGTCG  CCGCGGCAGC
      51  GGTTTGGGGC  GGATGGTCTT  AACTGAAGCC  CGAGCCGCAC  GTGCTTGATA
     101  TTACGGAAC  GGTCAAGCGC  GGC // .....
      //.. ATTTCGTTA  CGATTTGTG  CGAACCGGAT  ACGCCGATTA  AGGCGAAGCT
      51  CGACAGCGTC  GACCCCGGGC  TGACCACGAT  GTCGTCGGGC  GGTTACAACA
     101  GCAGTACGGA  TACGGCTTCC  AATGCGGTCT  ACTATTATGC  CCGTTCGTTT
     151  GTGCCGAATC  CCGACGGCAA  ACTCGCCACG  GGGATGACGA  CGCAGAATAC
     201  GGTGAAATC  GACGGCGTGA  AAAATGTGCT  GATTATTCCG  TCGCTGACCG
     251  TGAAAAATCG  CCGCGGCAAG  GCGTTGTGTC  GCGTGTGGG  TCGCGACGGC
     301  AAGGCGGCGG  AACGCGAAAT  CCGGACCGGT  ATGAGAGACA  GTATGAATAC
     351  CGAAGTAAAA  AGCGGGTTGA  AAGAGGGGGA  CAAAGTGGTC  ATCTCCGAAA
     401  TAACCGCCGC  CGAGCAACAG  GAAAGCGGCG  AACGCGCCCT  AGGCGGCCCG
     451  CCGCGCCGAT  AA

```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1  MAKMMKWAAV  AAVAAAAVWG  GWS.LKPEPH  VLDITETVRR  G.....
      51  .....
     101  .....
     151  .....
     201  .....
     251  PIKAKLDSVD  PGLTTMSSGG  YNSSTDASN  AVYYYARFV  PNPDKLATG
     301  MTTQNTVEID  GVKNVLIIPS  LTVKNRGGKA  FVRVLGADGK  AAEREIRTGM
     351  RDSMNTEVKS  GLKEGDKVVI  SEITAEQQE  SGERALGGPP  RR*

```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1  ..GTATCGGTCG  GCGCGCAGGC  ATCGGGGCAG  ATTAAGATAC  TTTATGTCAA
      51  ACTCGGGCAA  CAGGTAAAA  AGGGCGATTT  GATTGCGGAA  ATCAATTCGA
     101  CCTCGCAGAC  CAATACGCTC  AATACGGAAA  AATCCAAGTT  GGAACCGTAT
     151  CAGGCGAAGC  TGGTGTCCGC  ACAGATTGCA  TTGGGCAGCG  CGGAGAAGAA
     201  ATATAAGCGT  CAGGCGCGCT  TATGGAAGGA  AAACGCGACT  TCCAAAGAGG
     251  ATTTGGAAG  CGCGCAGGAT  GCGTTTGCCG  CCGCCAAAGC  CAATGTTGCC
     301  GAGCTGAAGG  CTTTAATCAG  ACAGAGCAAA  ATTTCCATCA  ATACCGCCGA
     351  GTCGGAATTG  GGCTACACGC  GCATTACCGC  AACGATGGAC  GGCACGGTGG
     401  TGGCGATTCT  CGTGGAAGAG  GGGCAGACTG  TGAACGCGGC  GCAGCTACG
     451  CCGACGATTG  TCCAATTGGC  GAATCTGGAT  ATGATGTTGA  ACAAATGCA
     501  GATTGCCGAG  GGCGATATTA  CCAAGGTGAA  GGCAGGGCAG  GATATTTCTG
     551  TTACGATTTT  GTCCGAACCG  GATACGCCGA  TTAAGGCGAA  GCTCGACAGC
     601  GTCGACCCCG  GGCTGACCAC  GATGTCGTCG  GGCAGTTACA  ACAGCAGTAC
     651  GGATACGGCT  TCCAATGCGG  TCTACTATTA  TGCCCGTTCT  TTTGTCCGGA
     701  ATCCGACGG  CAAACTCGCC  ACGGGGATGA  CGACGCAGAA  TACGTTGAA
     751  ATCGACGGCG  TGA AAAATGT  GCTGATTATT  CCGTCGCTGA  CCGTGA AAAA
     801  TCGCGGCGGC  AAGGCGTTG  TGCGCGTGT  GGGTGCGGAC  GGCAAGGCGG
     851  CGGAACGCGA  AATCCGACC  GGTATGAGAG  ACAGTATGAA  TACCGAAGTA
     901  AAAAGCGGGT  TGAAAGAGGG  GGACAAAGTG  GTCATCTCCG  AAATAACCGC
     951  CGCCGAGCAA  CAGGAAAGCG  GCGAACGCGC  CCTAGCGCGC  CCGCCGCGCC
    1001  GATAA

```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1  ..VSVGAQASQ  IKILYVKLGQ  QVKKGDIAE  INSTSQNTL  NTEKSKLETY
      51  QAKLVSAQIA  LGSAEKKYKR  QAALWKENAT  SKEDLESAQD  AFAAAKANVA
     101  ELKALIRQSK  ISINTAESEL  GYTRITATMD  GTVVAILVEE  GQTVNAAQST
     151  PTIVQLANLD  MMLNKMQIAE  GDITKVKAGQ  DISFTILSEP  DTPIKAKLDS
     201  VDPGLTTMSS  GGYNSSTDTA  SNAVYYYARS  FVPNPDGKLA  TGMTTQNTVE
     251  IDGVKNVLI  PSLTVKNRGG  KAFVRVLGAD  GKAAEREIRT  GMRDSMNTEV
     301  KSGLKEGDKV  VISEITAEQ  QESGERALGG  PPRR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

5	orf85.pep	MAKMMKWA	10	20	30	40	
	orf85a	MAKMMKWA	10	20	30	40	50 60
10	orf85.pep			80	90	100
	orf85a	TIVQLANL	210	220	230	240	250 260
15	orf85.pep	110	120	130	140	150	160
	orf85a	270	280	290	300	310	320
20	orf85.pep	170	180	190	200	210	220
	orf85a	330	340	350	360	370	380
30	orf85.pep	230					
	orf85a	390					

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

35	1	ATGGCAAAAA	TGATGAAATG	GGCGGCTGTT	GCGGCGGTCG	CGGCGGCAGC
	51	GGTTTGGGGC	GGATGGTCTT	ATCTGAAGCC	CGAGCCGCAG	GCTGCTTATA
40	101	TTACGGAAAC	GGTCAGGCGC	GGCGACATCA	GCCGGACGGT	TTCTGCAACA
	151	GGGGAGATTT	CGCCGTCCAA	CCTGGTATCG	GTCGGCGCGC	AGGCATCGGG
45	201	GCAGATTAAG	AAACTTTATG	TCAAACCTCG	GCAACAGGTT	AAAAAGGGCG
	251	ATTTGATTGC	GGAAATCAAT	TCGACCTCGC	AGACCAATAC	GCTCAATACG
50	301	GAAAAATCCA	AATTGGAAAC	GTATCAGGCG	AAGCTGGTGT	CGGCACAGAT
	351	TGCATTGGGC	AGCGCGGAGA	AGAAATATAA	GCGTCAGGCG	GCGTTGTGGA
55	401	AGGATGATGC	GACCGCTAAA	GAAGATTTGG	AAAGCGCACA	GGATGCGCTT
	451	GCCGCCGCCA	AAGCCAATGT	TGCCGAGCTG	AAGGCTCTAA	TCAGACAGAG
60	501	CAAAATTTCC	ATCAATACCG	CCGAGTCGGA	ATTGGGCTAC	ACGCGCATTG
	551	CCGCAACGAT	GGACGGCAGC	GTGGTGGCGA	TTCTCGTGGA	AGAGGGGCGC
65	601	ACTGTGAACG	CGGCGCAGTC	TACGCCGACG	ATTGTCCAAT	TGGCGAATCT
	651	GGATATGATG	TTGAACAAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
70	701	TGAAGGCGGG	GCAGGATATT	TCGTTTACGA	TTTGTCCGA	ACCGGATACG
	751	CCGATTAAGG	CGAAGCTCGA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
75	801	GTCGGGCGGC	TACAACAGCA	GTACGGATAC	GGCTTCCAAT	GCGGTCTACT
	851	ATTATGCCCC	TTCGTTTGTG	CCGAATCCGG	ACGGCAAAC	CGCCACGGGG
80	901	ATGACGACGC	AGAATACGGT	TGAAATCGAC	GGTGTGAAAA	ATGTGCTGAT
	951	TATTCCGTCG	CTGACCGTGA	AAAATCGCGG	CGGCAGGGCG	TTTGTGCGCG
85	1001	TGTTGGGTGC	AGACGGCAAG	GCGGCGGAAC	GCGAAATCCG	GACCGGTATG
	1051	AGAGACAGTA	TGAATACCGA	AGTAAAAAGC	GGGTTGAAAG	AGGGGGACAA
90	1101	AGTGGTCATC	TCCGAAATAA	CCGCCGCCGA	GCAGCAGGAA	AGCGGCGAAC
	1151	GCGCCCTAGG	CGGCCCGCCG	CGCCGATAA		

This encodes a protein having amino acid sequence <SEQ ID 770>:

60	1	MAKMMKWA	AAVAAAAVWG	GWSYLKPEPQ	AAYITETVRR	GDISRTVSAT
	51	GEISPSNLVS	VGAQASGQIK	KLYVKLGQOV	KKGDLIAEIN	STSQTNTLNT
65	101	EKSKLETYQA	KLVSQAIALG	SAEKYKRQA	ALWKDDATAK	EDLESAQDAL
	151	AAAKANVAEL	KALIRQSKIS	INTAESELGY	TRITATMDGT	VVAILVEEGQ
70	201	TVNAAQSTPT	IVQLANLDM	LNKMQIAEGD	ITKVKAGQDI	SFTILSEPDT
	251	PIKAKLDSVD	PGLTTMSSGG	YNSSTDASN	AVYYYARSFV	PNPDGKLATG
75	301	MTTQNTVEID	GVKNVLIIPS	LTVKNRGGRA	FVRVLGADGK	AAEREIRTGM

351 RDSMNTTEVK SGLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

5	orf85a.pep	30	40	50	60	70	80
	orf85-1	PQAA	YITETVRRGD	ISRTVSATGE	ISPSNLVSVGAQ	ASGQIKKLYVKLG	QQVKKGDLIAE
10	orf85a.pep	90	100	110	120	130	140
	orf85-1	INSTSQ	TNTLNTEKSK	LETYQAKLVSA	QIALGSAEKKYK	RQAALWKDDATA	KEDLESAQD
15	orf85a.pep	150	160	170	180	190	200
	orf85-1	ALAAKA	NVAELKALIR	QSKISINTAE	SELGYTRITAT	MDGTVVAILVEE	GQTVNAAQST
20	orf85a.pep	210	220	230	240	250	260
	orf85-1	PTIVQL	ANLDMMLNKM	QIAEGDITKV	KAGQDISFTIL	SEPDTPIKAKLD	SVDPGLTTMSS
25	orf85a.pep	270	280	290	300	310	320
	orf85-1	GGYNSST	DTASNAVYYY	ARSFVNPDPG	KLATGMTTQNT	VEIDGVKNVLI	IPSLTVKNRGG
30	orf85a.pep	330	340	350	360	370	380
	orf85-1	RAFVRVL	GADGKAAERE	IIRTGMRDSM	NTEVKSGLKEG	DKVVI SEITAAE	QQESGERALGG
35	orf85a.pep	390					
	orf85-1	PPRRX					
40	orf85a.pep	390					
	orf85-1	PPRRX					

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

Homology with a predicted ORF from *N. gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N. gonorrhoeae*:

50	ORF85	1	MAKMMKWAAVA	AAAAVWGGWS	.LKPEPHVLDITETVRRG.....	40
	ORF85ng	1	MAKMMKWAAVA	AAAAVWGGWS	YKPEPQAAYITEAVRRGD	ISRTVSAT 50
55	ORF85	201	TVNAAQSTPT	IVQLANLDMMLNKM	QIAEGDITKV	KAGQDISFTILSEPDT 250
	ORF85ng	201	TVNAAQSTPT	IVQLANLDMMLNKM	QIAEGDITKV	KAGQDISFTILSEPDT 250
60	ORF85	251	PIKAKLDSV	DPGLTTMSSGGYNS	STDASNAVYYYARS	FSFVNPDPGKLATG 300
	ORF85ng	251	PIKAKLDSV	DPGLTTMSSGGYNS	STDASNAVYYYARS	FSFVNPDPGKLATG 300
65	ORF85	301	MTTQNTVEI	DGVKNVLIIPSL	TVKNRGGKAFVRVL	GADGKAAEREI RTGM 350
	ORF85ng	301	MTTQNTVEI	DGVKNVLIIPSL	TVKNRGGKAFVRVL	GADGKAVEREIRTGM 350
	ORF85	152	RDSMNTTEVK	SGLKEGDKVVI	SEITAAEQQE	SGERALGGPPRR 393
	ORF85ng	351	KDSMNTTEVK	SGLKEGDKVVI	SEITAAEQQE	SGERALGGPPRR 393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1  ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CGGCGGCaac
51  GGTTCGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101  TTACGGAaac ggTCAGGCGC GGCATATCA GCCGGACGGT TTCCGCGACG
151  GgcgAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201  GCAGATTAAA AAGCTTTATG TCAAACTCGG GCAACAGGTC AAAAAGGGCG
251  ATTTGATTGC GGAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301  GAAAAATCCA AATTGGAAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
351  TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
401  AGGATGATGC GACCTCTAAA GAAGATTGG AAAGCGCGCA GGATGCGCTT
451  GCCGCGCCCA AAGCCAATGT TGCCGAGTTG AAGGCTTTAA TCAGACAGAG
501  CAAAATTTCC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATT
551  CCGCGACGAT GGACGGCAGC GTGGTGGCGA TTCCCGTGA AGAGGGGCG
601  ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
651  GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
701  TGAAGGCGGG GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
751  CCGATTAAAG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801  GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTATT
851  ATTATGCCCC TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCACGGGG
901  ATGACGACGC AGAATACGGT TGAATCGAC GGTGTGAAAA ATGTGTGTGT
951  TATTCGCTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACCGG
1001  TGTGGGTGTC GGACGGCAAG GCAGTGAAC GCGAAATCCG GACCGGTATG
1051  AAAGACAGTA TGAATACCGA AGTGAAGAG GGGTTGAAAG AGGGGGACAA
1101  AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
1151  GCGCCCTAGG CGCCCGCCG CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1  MAKMMKWA AVAAAVVWG GWSYLKPEPQ AAYITEAVRR GDISRTVSAT
51  GEISPSNLVS VQAQASGQIK KLYVKLGQV KKGDLIAEIN STTQNTIDM
101  EKSKLETYQA KLVSAQIALG SAEKKYKROA ALWKDDATSK EDLESAQDAL
151  AAAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201  TVNAAQSTPT IVQLANLDM LNMQIAEGD ITKVKAGQDI SFTILSEPDT
251  PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARSFV PNPDKLATG
301  MTTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTMG
351  KDSMNTEVKS GLKEGDKVVI SEITAEQQE SGERALGGPP RR*

```

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

30      40      50      60      70      80
orf85ng  PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDIAE
40      orf85-1  VSVGAQASGQIKLYVKLGQQVKKGDIAE
                        10      20      30

90      100     110     120     130     140
orf85ng  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKROAALWKDDATSKEDLESAQD
45      orf85-1  INSTSQNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKROAALWKENATSKEDLESAQD
                        40      50      60      70      80      90

150     160     170     180     190     200
orf85ng  ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVVAIPVEEGQTVNAAQST
50      orf85-1  AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
                        100     110     120     130     140     150

210     220     230     240     250     260
orf85ng  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
55      orf85-1  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
                        160     170     180     190     200     210

270     280     290     300     310     320
orf85ng  GGYNSSTDASNNAVYYYARSFVNPNDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
60      orf85-1  GGYNSSTDASNNAVYYYARSFVNPNDGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
                        220     230     240     250     260     270

330     340     350     360     370     380

```

-431-

```

orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf85-1      KAFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
                280      290      300      310      320      330

5          390
orf85ng      PPRRX
|||||
orf85-1      PPRRX

```

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from
membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia
coli] Length = 380

15 Score = 193 bits (485), Expect = 2e-48
Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

```

Query: 29  PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDIAE 88
          P   Y T  VR GD+ ++V ATG++      V VGAQ SGQ+K L V +G +VKK L+
Sbjct: 41  PVPTYQTLIVRPGDLQOSVLATGKLDALRKVDVGAQVSGQLKTL SVAIGDKVKKDQLLGV 100

```

```

20 Query: 89  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKYKROAALWKDDATSKEXXXXXX 148
          I+   N I  ++ L  +A+  A+  L  A  Y RQ  L  +  A S++
Sbjct: 101  IDPEQAENQIKEVEATLMELRAQRQQAELKLARVTYSRQRLAQT KAVSQDDLDTAAT 160

```

```

25 Query: 149  XXXXXXXXXXXXXXXXIROSKISINTAESDLGYTRITATMDGT VVAIPVEEGQTVNAAQST 208
          I++++ S++TA+++L YTRI A M G V I  +GQTV AAQ
Sbjct: 161  EMAVKQAQIGTIDAQIKRNQASLD TAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQA 220

```

```

30 Query: 209  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS 268
          P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++ V P
Sbjct: 221  PNILTLADMSAMLVKAQVSEADVIHLKPGQKAWFTVLGDPLTRYEQIKDVLP----- 273

```

```

35 Query: 269  GGYNSSTD TASNAVYYAR SFVPNP DGK LATGMTTONTVEIDGVKNVLLIPSLTVKNRGG 328
          + +  ++A++YYAR VPNP+G L  MT Q  +++ VKNVL IP  + +  G
Sbjct: 274  -----TPEKVND AIFYYARFEV PNPNGLLRLDMTAQVHIQLTDVKNVLTIPLSALGDPVG 328

```

```

Query: 329  KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS E 372
          +V L  +G+  ERE+  G ++  + E+  GL+ GD+VVI E
Sbjct: 329  DNRYKVKLLRNGETREREVTIGARNDT DVEIVKGLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein
45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

50      1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
      51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
     101  CGGTTGTCGG CAATACCTCG CACCCTACCT ACTATAGAGA CATACGCAGG
     151  GGCAAACTGT ATGCGGAAGc CAAATTCGCC GACgGcAGCG TAACTTACGG
     201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAGGCT ATGGATTGT

```

-432-

5
251 TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCGGGG
301 CTGAAAATCA CCAACGGCAA AAAACTTTAT TCCGTCGGCG GTTGAATAA
351 GGCGGGTACA GGAAAATACA GCATAGGCGG CGTGGAAACC GAAGTCGTCA
401 AATATCGGGT GCGGCGCGGC GACGATGCGG TAATGTATTT CTTCGCACCG
451 TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAC
501 CTATACGCTG AAACCTCAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC
551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

10
1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR
51 GKLYAEAKFA DGSVTYKGAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP
151 SLNNIPAQIG YTDGKTYTL KLKSVQINGQ AAKP*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

15
1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT
101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201 TTTCGAGTCC GGCGGTACGG TTGTCCGCAA TACCTGCAC CCTACCTACT
20
251 ATAGAGACAT ACGCAGGGGC AAACGTATG CGGAAGCCAA ATTCCGCCGAC
301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
351 CAAGGCTATG GATTGTTCG CGCTTGCTG GCAGTTGGCG GCAATGACG
401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAA ACTTTATTCC
451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
501 GGAACCGAA GTCGTCAAAT ATCGGGTGGC GCGCGGCGAC GATGCGGTAA
25
551 TGTATTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601 ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651 CGGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

30
1 MMKTFKNIFS AAILSALPC AYAAGLPQSA VLHYSYSGYI PATMTFERSG
51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
101 GSVTYKGAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

40
orf120.pep IPATMTFERSGNAYKIVSTIKVPLYNIRFE
orf120a SAAILSAALPCAYAAGLPXSAVLHYSYSGYIPATXXXXXXXXNAXKIVSTIKVPLYNIRFE
10 20 30 40 50 60
45
orf120.pep SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAGESKTEQSPKAMDFTLAWQL
orf120a SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAXXXXXXQSPKAMDFTLAWQL
70 80 90 100 110 120
50
orf120.pep AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP
orf120a AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP
130 140 150 160 170 180
55

-433-

```

              160      170      180
orf120.pep  SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
              |||||
orf120a     SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
5           190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

      1  ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
     51  CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
    101  ATTCCGGCAG CTACGGCATT CCCGCCACNA NNANNTNNGN ACNNGNGNCG
    151  AATGCTTNCA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
    201  TTTTCGAGTCC GGCGGTACGG TTGTTCGGCAA TACCCTGCAC CCTACCTACT
    251  ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
    301  GGCAGCGTAA CCTACGGCAA AGCGGNNNNN ANCNNNNNNG NGCAAAGCCC
    351  CAAGGCTATG GATTGTTCG CGCTTGCTG GCAGTTGGCG GCAAATGACG
    401  CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
    451  GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
    501  GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
    551  TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
    601  ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
    651  CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

      1  MMKTFKNIFS AAILSAAALPC AYAAGLPXSA VLHYSYSGSYGI PATXXXXXXXX
     51  NAXKIVSTIK VPLYNIRFES GGTVVGNLTH PTYYRDIRRG KLYAEAKFAD
    101  GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
    151  VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
    201  TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
orf120a.pep  MMKTFKNIFS AAILSAAALPC AYAAGLPXSA VLHYSYSGSYGI PATXXXXXXXX NAXKIVSTIK
              |||||
orf120-1     MMKTFKNIFS AAILSAAALPC AYAAGLPQSA VLHYSYSGSYGI PATMTFERSGNAYKIVSTIK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf120a.pep  VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAM
              |||||
orf120-1     VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
              70      80      90      100     110     120

              130     140     150     160     170     180
orf120a.pep  DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
              |||||
orf120-1     DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
              130     140     150     160     170     180

              190     200     210     220
orf120a.pep  DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
              |||||
orf120-1     DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
              190     200     210     220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55  orf120.pep  IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
     orf120ng  SAAILSAALPCAYAARLPQSAVLHYSYSGSYGI PATMTFERSGNAYKIVSTIKVPLYNIRFE 69

60  orf120.pep  SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 90
     orf120ng  SGGTVVGNLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 129

```

5
 orf120.pep AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP 150
 orf120ng AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDTVTYFFAP 189
 orf120.pep SLNNIPAIQIGYTDGKTYTLKLKSVQINGQAAKP 184
 orf120ng SLNNIPAIQIGYTDGKTYTLKLKSVQINGQAAKP 223

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

10
 1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
 51 CCTGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT
 101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
 201 TTTCAATCC GGCGGTACGG TTGTCGGCAA TACCCTGCAC CCTGCCTACT
 15 251 ATAAAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
 301 GGCAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
 351 CAAGGCTATG GATTTGTTCA CGCTTGCCTG GCAGTTGGCG GCAAATGACG
 401 CGAAACTCCC CCCGGGTCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
 451 GTCGGCGGCC TGAATAAGGC GGGTACGGGA AAATACAGCA TaggCGGCGT
 20 501 GGAAACCGAA GTCGTCAAT ATCGGGTGCG GCGCGGCGAC GATACGGTAA
 551 CGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
 601 ACCGACGACG GCAAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA
 651 CGGACAGGCC GCCAAACCGT AA

This encodes a protein having amino acid sequence <SEQ ID 780>:

25
 1 MMKTFKNIFS AAILSAALPC AYAARLPQSA VLHYSYSGYGI PATMTFERSG
 51 NAYKIVSTIK VPLYNIRFES GGTVVGNLTH PAYYKDIRRG KLYAEAKFAD
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAIQIGY
 201 TDDGKTYTLK LKSVQINGQA AKP*

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

35
 orf120-1.pep 10 20 30 40 50 60
 MMKTFKNIFS AAILSAALPCAYAAGLPQSAVLHYSYSGYGI PATMTFERSGNAYKIVSTIK
 orf120ng MMKTFKNIFS AAILSAALPCAYAARLPQSAVLHYSYSGYGI PATMTFERSGNAYKIVSTIK
 40
 orf120-1.pep 70 80 90 100 110 120
 VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADG SVTYGKAGESKTEQSPKAM
 orf120ng VPLYNIRFESGGTVVGNLHPAYYKDIRRGKLYAEAKFADG SVTYGKAGESKTEQSPKAM
 45
 orf120-1.pep 130 140 150 160 170 180
 DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
 orf120ng DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
 50
 orf120-1.pep 190 200 210 220
 DAVMYFFAPSLNNIPAIQIGYTDGKTYTLKLKSVQINGQAAKPX
 orf120ng DTVTYFFAPSLNNIPAIQIGYTDGKTYTLKLKSVQINGQAAKPX
 190 200 210 220

55 This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

-435-

```

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
101 CTCCGTTTGC GGTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGCTGTG
5 GATGGTGTTT TCCTTGATTT TGTGTGTGGC ATTATTGTTG ATTATCGTCC
251 CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCCAATTA
301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
10 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
451 AGGCAGGGCG GCAATATT..

```

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

```

1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF>NNLASRLPQL
15 IGFQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
151 RQGGNI..

```

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

```

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
51 GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
101 CTCCGTTTGC GGTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
20 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGCTGTG
201 GATGGTGTTT TCCTTGATTT TGTGTGTGGC ATTATTGTTG ATTATCGTCC
251 CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCCAATTA
301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
25 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC
501 CTTGCTGCTT TACTATTTC TGTGGATTG GCAGCGGTGG TCGTGCGGCA
551 TTGCCAAACT GGTTCCGAGG CGTTTTGCCG GTGCTTATAC GCGCATTACA
601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT
30 AATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGGTG CTGGTCGGGC
701 TGGATTCGGG GTTTCGCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTC
751 CCTTATCTCG GGGCGTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCCTT
801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG
851 CCGTAGGACA GTTCTCTGAA AGTTTTTTCA TTACGCCGAA AATCGTGGGA
35 GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGCGGTT
951 CGGGCAGCTG ATGGCTTTTG TCGGAATGTT GGCGGGATTG CCTTTGGCCG
1001 CCGTAACCTT GCTCTTGCTT CGCGAGGGCG TGCAGAAATA TTTTGGCCGC
1051 AGTTTTTACC GGGGCAGGTA G

```

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

```

40 1 MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF>NNLASRLPQL
101 IGFQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
151 RQGGNIVSSI GNLLLLPLLL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT
201 GNINEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMLAGILVFE
45 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG
301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAATLVLL REGVQKYFAG
351 SFYRGR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

```

10 20 30 40 50 60
orf121.pep MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR
55 orf121a MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR
10 20 30 40 50 60
70 80 90 100 110 120
orf121.pep ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV

```

-436-

```

      |||
orfl21a  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
              70      80      90      100      110      120
5
      130      140      150
orfl21.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
      |||
orfl21a  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
              130      140      150      160      170      180
10
orfl21a  SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
              190      200      210      220      230      240

```

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

```

1  ATGTATCGGA  GGAAAGGGCG  GGGCATCAAG  CCGTGGATGG  ATGCCGGTGC
15  51  GCGCTTTGCC  GCCTTGGTCT  GGCTGGTTTT  CGCGCTCGGC  GATACTTTGA
      101  CTCCGTTTGC  GGTGCGGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
      151  GAATGGTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGTCTGT
      201  GATGGTGTTC  TCCTTGATTT  TGTGTTGGC  ATTATTGTTG  ATTATTGTCC
      251  CTATGCTGGT  CGGGCAGTTC  AACAATTTGG  CATCGCGCCT  GCCCCAATTA
20  301  ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAAA  ATACAATCGG
      351  CCGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTACAGGCG
      401  ATACGGGCGA  GTTAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTGTATG
      451  AGGCAGGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCTTCC
      501  CTTGCTGCTT  TACTATTTCC  TGCTGGATTG  GCAGCGGTGG  TCGTGC GGCA
25  551  TTGCCAAACT  GGTTCGAGG  CGTTTTCGG  GTGCTTATAC  GCGCATTACA
      601  GGCAATTTGA  ACGAGGTATT  GGGCGAATTT  TTGCGCGGGC  AGCTTCTGGT
      651  GATGCTGATT  ATGGGTTTGG  TTTACGGCTT  GGGGTGTTGG  CTGGTCGGGC
      701  TGGATTCGGG  GTTGTCAATC  GGTATGGTTG  CCGGTATTTT  GGTTTTGTGT
      751  CCCTATTTGG  GCGCGTTTAC  AGGACTGCTG  CTGGCAACCG  TCGCCGCCTT
30  801  GCTCCAGTTC  GGTTCGTGGA  ACGGCATCTT  GGCTGTTTGG  GCGGTTTGTG
      851  CCGTAGGACA  GTTCTCGAA  AGTTTTTCA  TTACGCCGAA  AATCGTGGGA
      901  GACCGTATCG  GCCTGTCGCC  GTTTTGGGTT  ATCTTTTCGC  TGATGGCGTT
      951  CGGGCAGCTG  ATGGGCTTTG  TCGGAATGTT  GGCCGGATTG  CCTTTGGCCG
35  1001  CCGTAACCTT  GGTCTTGCTT  CGCGAGGGCG  TGCAGAAATA  TTTTGCCGGC
      1051  AGTTTTTACC  GGGGCAGGTA  G

```

This encodes a protein having amino acid sequence <SEQ ID 786>:

```

1  MYRRKGRGIK  PWDAGAAFA  ALVVLVFALG  DTLTPFAVAA  VLAYVLDPLV
40  51  EWLQKKGLNR  ASASMSVMVF  SLILLALLL  IIIVPMLVGQF  NNLASRLPQL
      101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELSN  ALKAWFPVLM
      151  RQGGNIVSSI  GNLLPLLL  YYFLDWQRW  SCGIAKLVPR  RFAGAYTRIT
      201  GNLNEVLGEF  LRGQLLVMLI  MGLVYGLGLV  LVGLDSGFAI  GMVAGILVVF
      251  PYLGAFTGLL  LATVAALLQF  GSWNGILAVW  AVFAVGQFLE  SFFITPKIVG
      301  DRIGLSPFWV  IFSLMAFGQL  MGFVGMLAGL  PLAATVLVLL  REGVQKYFAG
351  SFYRGR*

```

ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```

      10      20      30      40      50      60
orfl21a.pep MYRRKGRGIKPWMDAGAAFAALVVLVFALGDTLTPFAVAAYLAYVLDPLVEWLQKKGLNR
      |||
50  orfl21-1  MYRRKGRGIKPWMDAGAAFAALVVLVFALGDTLTPFAVAAYLAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60
      70      80      90      100      110      120
orfl21a.pep ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      |||
55  orfl21-1  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      70      80      90      100      110      120
      130      140      150      160      170      180
orfl21a.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      |||
60  orfl21-1  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      130      140      150      160      170      180
      190      200      210      220      230      240
65  orfl21a.pep SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI

```

10

15

20

25

30

35

40

45

50

55

60

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```

      1 MYRRKGRGIK PWMGAGAAFA ALVWLIVYALG DTLTPFAVAA VLAYVLDPLV
      51 EWLQKKGLNR ASASMSVMVF SLILLLALLL IIVPMLVGQF>NNLASRLPQL
      101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
5      151 KQGGNIVSSI GNLLLPPLL YYFLLDWQRW SCGIAKLVPR REAGAYTRIT
      201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLM LVGLDSGFAI GMVAGILVFV
      251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
      301 DRIGLSPFWV IFSLMAFGEL MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
      351 SFYRGR*

```

10 ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

```

      10      20      30      40      50      60
orff121-1.pep MYRRKGRGIKPWMGAGAAFAALVWLIVFALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR
      15      10      20      30      40      50      60
orff121ng-1 MYRRKGRGIKPWMGAGAAFAALVWLIVYALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR

      70      80      90      100     110     120
orff121-1.pep ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      20      70      80      90      100     110     120
orff121ng-1 ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV

      130     140     150     160     170     180
orff121-1.pep EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLPLLLYYFLLDWQRW
      25      130     140     150     160     170     180
orff121ng-1 EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSSIGNLLLPLLLYYFLLDWQRW

      190     200     210     220     230     240
orff121-1.pep SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
      30      190     200     210     220     230     240
orff121ng-1 SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI

      250     260     270     280     290     300
orff121-1.pep GMLAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
      35      250     260     270     280     290     300
orff121ng-1 GMVAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG

      310     320     330     340     350
orff121-1.pep DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
      40      310     320     330     340     350
orff121ng-1 DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX

```

In addition, ORF121ng-1 shows homology to a permease from *H.influenzae*:

```

sp|P43969|PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
Score = 69.9 bits (168), Expect = 2e-11
Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)

50 Query: 26 VYALGDTLTPFAVAALVAYVLDPLVEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXXXXVP 84
      +Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
Sbjct: 32 IYFFGDLIAPLLIALVLSYLLEIPINFLNQYLCPRMLATILIFGSFIGLAFFVLVLP 91

55 Query: 85 MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK 143
      ML Q +L S LP + N WL N Y E ID + + + F + ++ +
Sbjct: 92 MLWNQTISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFE 147
Query: 144 AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXXDWQRWSCGIAKLVPRRFAGAYTRITGNL 203

60 Sbjct: 148 SAVKLSLASIMNLVSLGIYAFVLPMFFMLKDKSELLQGVSRFLPKNRNLAFXRWK-EM 206

Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXDSGFAIGMVAGILVFVPYXXXXXXXXXXXXX 263
      + + ++ G+ + + G+ V VPY
65 Sbjct: 207 QQQISNYIHGKLEILIVTLITYIIFLIFGLNYPLLAFVAVGLSVLPYIGAVIVTIPVA 266

```

-439-

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIISVLIFGGLWGF 326

5 Query: 324 VGMLAGLPLAAVTLVLL 340
 G+ +PLA + ++
 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

```

15 1  .ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
    51  TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
    101 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
    151 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
    201 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
    251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTCTGTGTC
20 301 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
    351 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACCTCT
    401 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
    451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
    501 CGAGCAGCCC TTTTTCAAAT GGGATTCAA CTCCGCCAAG TATCAG..

```

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

```

    1  .TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR
    51  LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRRECGFLC
    101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCRT
    151 EQRVGNVQQQ RIGIGVSEQP FFKWDFNSAK YQ..

```

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

```

    1  ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGGGAAG TAGATACCGC
    51  GCCTTTGATT TTTTGGCCGC TCTTACCAA GGCTTCGATG AAAAAGTTGA
    101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTCGGG TACGAATTCTG
    151 ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTGTG TCGTCATATT
35 201 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
    251 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
    301 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
    351 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
    401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTTGG GTTCTGTGTC
40 451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
    501 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACCTCT
    551 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
    601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
    651 CGAGCAGCCC TTTTTCAAAT GGGATTCAA CTCCGCCAAG TATCAGCTTT
45 701 CTGCTTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT
    751 CGTCATCGTT TGTGTTCTCT A

```

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

```

    1  ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMPVPVPM PIYSFSGTNS
    51  TAFSAAMRLS SSCVVIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR
50 101  LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRREFGFLC
    151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCRT
    201 EQRVGNVQQQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDDTV
    251 RHRLCS*

```

Computer analysis of this amino acid sequence gave the following results:

5	orf122.pep				10	20	30
					TAFSAALRLSPSXLVIFLSFGKPYQQTAAI		
	orf122a	FLPLLPKASMKKLMVEPVMPMPYSFSGTNSTAFSAAMRLSSSCVVI			: :		
		30 40 50 60 70 80					
10		40 50 60 70 80 90					
	orf122.pep	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR					
	orf122a	LTFFXTSCPPRSNPYQQYRRLRLYAFHAPEITEFFVGFAFXVDARNVYAQIGGDVGTHLR					
		90 100 110 120 130 140					
15		100 110 120 130 140 150					
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT					
	orf122a	NMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT					
20		150 160 170 180 190 200					
		160 170 180					
	orf122.pep	EQRVGNGVQQRIGIGVSEQPFFKWDFNSAKYQ					
	orf122a	EQRVGNGVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLVDIVALSDTDVVRHRLCSX					
25		210 220 230 240 250					

	1	ATATCATATT	GGGCAAGCAG	TTCACTGGAT	TTTTTGGGAAG	TAGATACCGC
30	51	GCCTTTGATT	TTTTTGCCGC	TCTTACCCAA	GGCTTCGATG	AAAAAGTTGA
	101	TGGTCGAACC	GGTACCGATG	CCGATGTATT	CGTTTTCCGG	TACGAATTCC
	151	ACTGCTTTT	CGCGGCGAT	GCGCTTGAGT	TCGTCTTG	TCGTATATT
	201	TTTGTCTTT	GGGAAACCGT	ATCAACAAAC	AGCCCGCATC	TTAACATTTT
	251	TTNNNACGTC	CTGCCGCGC	CGTTCAAATC	CTTACCGAGT	ATACCGCCGC
35	301	CTGCGACTCT	ATGCCTTCCA	TGCGCCCGAG	ATAACCGAGT	TTTTCTGTGG
	351	TTTGGCTTT	GANGTTGACG	CACGAAATGT	CTATGCCCAA	ATCGGCGCGC
	401	ATGTTGGCAC	GCATTTGCGG	AATATGCGGC	GCGAGTTTGG	GTTTCTGTGC
	451	AATCACGCTC	GTATCGACAT	TGACCGCCTG	CCAACCCCTG	GCCTGAACCG
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGCG	TGTCGCCATC	TTTGAACCTC
40	551	GCGGCGGTGT	CGGGGAAATG	GCTGCCGATA	TGCCCAAAAC	CTGCCGACCC
	601	GAGCAGCGCG	TCGGTAACGG	CGTGCAGCAG	CGCATCGGCA	TCGGAGTGTC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCCTTCG	TCAGTTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATGTT
	751	CGTCACTGTT	TGTGTTCCCTG	A		

45	1	ISYWASSSLD	FLEVDTAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQOTAAI	LTFFXTSCPP	RSNPYQOYRR
	101	LRLYAFHAFE	ITEFFVGFAF	XVDARNVYAQ	IGGDVTHLPR	NMRREFGFLC
	151	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGEM	AADIAQTCRT
	201	EQRVGNVQVQ	RIGIGVSEQP	FFKWFENSAK	YQLSAFGQLV	DIVALSDTDV
50	251	RHRLCS*				

```

10      20      30      40      50      60
orf122a.pep  ISYWASSSLDFLEVDTAFLIFLPLLPKASMKKLMVEPVMPMPYSFSGTNSTAFSAAMRLS
55          |||||:|||||
orf122-1     ISYWASSSPDFLEVDTAFLIFLPLLPKASMKKLMVEPVMPPIYSFSGTNSTAFSAAMRLS
          10      20      30      40      50      60

          70      80      90     100     110     120
orf122a.pep  SSCVVIFLSFGKPYQQTAAILTFFXTSCPPRSNPYQQYRRLRLYAFHAPETEFFVGFAF
60          |||||:|||||
orf122-1     SSCVVIFLSFGKPYQQTAAILTEFTCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF

```

-441-

		70	80	90	100	110	120
5	orf122a.pep	130	140	150	160	170	180
	orf122-1	130	140	150	160	170	180
10	orf122a.pep	190	200	210	220	230	240
	orf122-1	190	200	210	220	230	240
15	orf122a.pep	250					
	orf122-1	250					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI	80
30	orf122.pep	LTFECTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
	orf122ng	LTFECTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
35	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
	orf122ng	NVRCEFGFLCNHGRIDIDHPLTLRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
	orf122.pep	EQRVGNVGQQRIGIGVSEQPFFKWFDFNSAKYQ	182
	orf122ng	EQRVGNVGQQRVGIRMPEQPFFKWFDFNSAKYQLSAFGQLVDIVALSDDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTCGCCGAT	TTTTTGGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTTGCCCAA	GGCTTCGATG	AAGAAATTGa
	101	tgGTCGAACC	GgtaCCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCG
	151	ACTGCTTTT	CGGCGCGCAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
45	201	TTTAtccttt	gGGAaacct	atcaAcaAAc	agccgccatC	TTAACATTTT
	251	TTTGACGtc	ctggccgcgc	cgttcaAATc	cgtaccaGca	ataccgccgc
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GATatTGACG	CACGAAATAT	CGatacCCAa	atcgccgcGCG
	401	ATGTTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
50	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAATCT
	551	GCGGCGGTGT	CGGGAAAATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
	601	GAGCAGCgcg	tcggtaaCGG	CGTGCAGCAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
55	701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
	751	CGTCATCGTT	TGTGTTCTCG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

60	1	MSYRASSPD	FLEVETAPLI	FLPLLKASM	KKLMVEPVEM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVIFLSF	GKPYQQTAAI	LTFECTSWPP	RSNPYQQYRR
	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDHL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
	201	EQRVGNVGQQ	RVGIRMPEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDDI
	251	RHRLCS*				

-442-

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVDTAPLI	FLPLLPKASMKKLMVEPV	MPPIYSFSGTNSTAFSA	MRLS		
5	orf122ng	MSYRASSSPDFLEVETAPLI	FLPLLPKASMKKLMVEPV	MPMYSFSGTNSTAFSA	MRLS		
		10	20	30	40	50	60
	orf122-1.pep	SSCVVIFLSFGKPYQQTAA	ILTFCTSCPPRSNAYQQY	RRLRLYAFHPPEIAEFF	VGF		
10	orf122ng	SSCVVIFLSFGKPYQQTAA	ILTFCTSWPPRSNPYQQY	RRLRLYAFHPPEIAEFF	VGF		
		70	80	90	100	110	120
	orf122-1.pep	DVDARNVYAQIGGDVGTH	LRNVREFGFLCNHGRID	IDLPTLRLNALIRRTQ	KDA		
15	orf122ng	DIDARNIDTQIGGDVGTH	LRNVRCEFGFLCNHGRID	IDLPTLRLNALIRRTQ	KDA		
		130	140	150	160	170	180
	orf122-1.pep	FELCGGVGEMAADIAQTC	RTEQRVGNVQQRIGIGV	SEQPFKWFDFNSAKYQ	LSA		
20	orf122ng	FELCGGVGKMAADVAQTC	RTEQRVGNVQQRVGIRM	PEQPFKWFDFNSAKYQ	LSA		
		190	200	210	220	230	240
	orf122-1.pep	DIVALSDTDVHRRLCSX					
25	orf122ng	DIVALSDTDIRHRLCSX					
		250					
30							

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGGCGCGA	GTGCGAACAA	CATTTCCGCG	CGTTTTGCGG	AAACACCCGT
	51	CGCTGTCAGC	GTTACCCTGA	TCGGCACGGT	ACTTGCCGTC	ATGCTGCCCC
	101	TTACCGAATA	TGAAACTTC	CTGCTGCTTA	TCGGCTCGGT	ATTTGCGCCG
	151	ATGGGGCGGA	TTTGATTGC	CGACTTTTTC	GTCTTGAAAC	GGCGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	..AGASANNISA	RFAETPVAVS	VTLIGTVLAV	MLPVTEYENF	LLLIGSVFAP
	51	MGGFDCRLFR	LETA*			

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCTCCGCCA	TCGGGCTGAT
	51	TTGGTTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACGCTGC
45	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CGGCTCTACT	TTTGGGTGAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTCGGGCG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGGAAA	GCGTGCGCCT	GTCGTTCGGC	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATT	ACGCCGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
50	351	GTGGACGGC	GAATCTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
	401	TTGTGCTGTG	GCTGGTTTTC	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GTTGGCGGTT	CTGTGGCTGA	GTGCCGAAGT
	501	CTTTTCCACG	GCAGGCAGCA	CCGCCGCACA	GGTTTCAGAC	GGCATGAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTGA	TGCCGCTTTC	CTGGCTGCCG
55	601	CTTGCCGCCG	ACTACACGCG	CCACGCGCGC	CGCCCGTTTG	CGGCAACCCCT
	651	GACGGCAACG	CTCGCCTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTGG
	701	GTTTGGCAGC	GGCGTTGTTC	ACCGGAGAAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GGCAGGCATT	TTGGCGGTGC	TCCTCTCCAC

-443-

5
 801 CGTTACCACA ACGTTTCTCG ATGCCTATTC CGCCGGCGCG AGTGCGAACA
 851 ACATTTCCGC GCGTTTTCG GAAACACCCG TCGCTGTCCG CGTTACCCTG
 901 ATCGGCACGG TACTTGCCGT CATGCTGCCC GTTACCGAAT ATGAAAACCT
 951 CCTGCTGCTT ATCGGCTCGG TATTTGCGCC GATGGCGGCG GTTTTGATTG
 1001 CCGACTTTT CGTCTGAAA CGGCGTGAGG AGATTGAAGG CTTTGACTTT
 1051 GCCGGACTGG TTCTGTGGCT TGCGGGCTTC ATCCTCTACC GCTTCCTGCT
 1101 CTCGTCCGGC TGGGAAAGCA GCATCGGTCT GACCGCCCCC GTAATGTCTG
 1151 CCGTTGCCAT TGCCACCGTA TCGGTACGCC TTTTCTTTAA AAAAACCCAA
 1201 TCTTTACAAA GGAACCCGTC ATGA

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

15
 1 MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSSA NMLQLAGWTA
 101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
 151 VSMLLMLLAV LWLSAEVFST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLP
 201 LAADYTRHAR RPFATLTAT LAYTLTGCM YALGLAAALF TGETDVAKIL
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETPVAVGVTL
 301 IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFDF
 351 AGLVLWLAGF ILYRFLSSG WESSIGLTAP VMSAVAIATV SVRLFEEKTO
 401 SLQRNPS*

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

25
 orf125.pep AGASANNISARFAETPVAVSVTLLIGTVLAV
 orf125a KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAVVGTLLAV
 250 260 270 280 290 300
 30
 orf125.pep MLPVTEYENFLLIGSVFAPMGGFDCRLFRLETAX
 orf125a LLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG
 310 320 330 340

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

40
 1 ATGTCGGGCA ATGCCTCCTC TCNTTCATCT TCCGCCGCCA TCGGGCTGAT
 51 TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACACTGC
 101 TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CNGCTCTGCT TTTGGGTGCT
 151 GCCGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC
 201 CGGACNCANC TCGATGGAAA GCGTGCGCCT GTCGTTCCGGC AAACGCGGTT
 251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
 301 GTGATGATT ACGCCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
 351 GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GGCGCGCTGA
 401 TTGTGCTGTG GCTGGTTTTC GGCGCACGCA AAACAGGCGG GCTGAAACC
 451 GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAANT
 501 NTTTTCCACG GCAGGCAGCA CCGCCGCANN GGTNNCAGAC GGCATGAGTT
 551 TCGGAACGGC AGTCGAGCTG TCCGCCGTNA TGCCGCTTTC TTGGCTGCCG
 601 CTGGCCGCCG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCCCT
 651 GACGGCAACG CTCGCCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG
 701 GTTTGGCAGC GCGCTTGTTC ACCGGAGAAA CCGACGTGGC AAAAATCCTG
 751 CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTCG TCCTGTGCGAC
 801 CGTTACCACC ACTTTTCTCG ATGCNTACTC CGCCGCGGTA AGTGCCAACA
 851 ATATTTCCGC CAAACTTTCG GAAATACCNA TCGCCGTTGC CGTCGCGGTT
 901 GTCGGCACAC TGCTTGCCGT CCTCTGCCC GTTACCGAAT ATGAAAACCT
 951 CCTGCTGCTT ATCGGCTCGG TATTTGCGCC GATGGCGGCG GTTTTGATTG
 1001 CCGACTTTTT CGTCTTGAAG CGGCGTGAGG AGATTGAAGG C..

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSSA NMLQLAGWTA

-444-

101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP
 201 LAADYTRHAR RPFATLTAT LAYTLTGCM YALGLAAALF TGETDVAKIL
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPIAVAVAV
 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSAAIGLIWFGAAVSIAEISTGTLAPLGWQRLAALLLGHAVGGALFFAA					
	orf125-1	MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLAPLGWQRLAALLLGHAVGGALFFAA					
		10	20	30	40	50	60
15	orf125a.pep	AYIGALTGXSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
	orf125-1	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
		70	80	90	100	110	120
20	orf125a.pep	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEXFSTAGSTAAXVXD					
	orf125-1	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSSTAGSTAQVSD					
		130	140	150	160	170	180
25	orf125a.pep	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAAALF					
	orf125-1	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAAALF					
		190	200	210	220	230	240
30	orf125a.pep	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAV					
	orf125-1	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVTL					
		250	260	270	280	290	300
35	orf125a.pep	VGTLLAVLLPVTEYENFLLLIGSVFAPMAAVLIADFFVLKRREEIEG					
	orf125-1	IGTVLAVMLPVTEYENFLLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAGLVWLWLAGF					
		310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
50	orf125.pep	MLPVTEYENFLLLIGSVFAPM-GGFDCLFRLETA	64
	orf125ng	MLPVTEYKNFLLLIRSVFGPMAGGFDCLFCLKTA	343

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
 101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
 151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAMPLSWL
 201 PLAADYTRQA RRPFAATLTA TLAYTLTGCM MYALGLAAL FTGETDVAKI
 251 LLGAGLGITG ILAVVLSTVT TTFLDTYSAG ASANNISARF AEIPVAVGV
 301 LIRTVLAVML PVTEYKNFLL LIRSVFGPMA GGFDCLFCL KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

      1 ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
    51 TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
  101 TCGCCCCCTT GGGCTGGCAG CGCGGTCTGG CCGCCCTGCT TTTGGGTCAT
    5 GCGGTCGGCG GCGCGCTGTT TTTTGCGGCG GCGTATATCG GCGCACTGAC
  201 CGGACGCAGC TCGATGGAAA GTGTGCGCCT GTCGTTCCGG AAATGCGGTT
  251 CAGTGCTGTT TTCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
  301 GTGATGATTT ACGTCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGT
  351 CTGGGACGGC GAATCCTTTG TCTGGTGGGC ATTGGCAAAC GGCGCACTGA
  10 401 TCGTGCTGTG GCTGGTTTTT GCGGCACGCA GAACGGGCGG GCTGAAAACC
  451 GTTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGGTTGA GCGTCGAAGT
  501 GTTCGCTTCG TCCGGCACAA ACGCCGCGCC CGCCGTTTCA GACGGCATGA
  551 CCTTCGGAAC GGCAGTCGAA CTGTCCGCCG TCATGCCGCT TTCCTGGCTG
  601 CCGCTGGCCG CCGACTACAC GCGCCAAGCA CGCCGCCCGT TTGCGGCAAC
  15 651 CCGTACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCCT
  701 TGGGTTTGGC GCGGCTCTG TTTACCGGAG AAACCGACGT GCGGAAAATC
  751 CTGTTGGGCG CCGGCTTGGG CATAACGGGC ATTCTGGCAG TCGCTCTCTC
  801 CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTGCGA
  851 ACAACATTTT CCGCGGTTTT GCGGAAATAC CCGTCGCTGT CCGCGTTACC
  20 901 CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
  951 CTTCTGCTG CTTATCGGCT CGGTATTTGC GCCGATGGCG GCGGTTTTGA
  1001 TTGCCGACTT TTTCGTCTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
  1051 TTTGCCGGAC TGCTTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT
  1101 GCTCTCGTCC GGTGGGAAA GCAGCATCGG TGTGACCGCC CCCGTTATGT
  25 1151 CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
  1201 CAATCTTTAC AAAGGAACCC GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

      1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
    51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
  30 101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
  151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
  201 PLAADYTRQA RRPFAATLTA TLAYTLTGCV MYALGLAAAL FTGETDVAKI
  251 LLGAGLGITG ILAVVLSTVT TTFDLTYSAG ASANNISARF AEIPVAVGVT
  301 LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
  35 351 FAGLVLWLAG FILYRFLLS GWESSIGLTA PVMSAVAIAT VSVRLFEEKT
  401 QSLQRNPS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

      10      20      30      40      50      60
  40 orf125-1.pep  MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLAPLGWQ RGLAALLLGHAVGGALFFAA
      |||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     MSGNASSPSSSAAIGLVWFGAAVSIAEISTGTLAPLGWQ RGLAALLLGHAVGGALFFAA
      10      20      30      40      50      60

      70      80      90     100     110     120
  45 orf125-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
      |||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
      70      80      90     100     110     120

      130     140     150     160     170     179
  50 orf125-1.pep  ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAAQ-VS
      |||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSGTNAAPAVS
      130     140     150     160     170     180

      180     190     200     210     220     230     239
  55 orf125-1.pep  DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCVWYALGLAAAL
      |||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCVWYALGLAAAL
      190     200     210     220     230     240

      240     250     260     270     280     290     299
  60 orf125-1.pep  FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDTYSAGASANNISARFAETPVAVGVT
      |||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVT
      250     260     270     280     290     300

```

-446-

		300	310	320	330	340	350	359
	orf125-1.pep	LIGTVLAVMLPVT	EYENFLL	LIGSVFAPMA	AVLIADFFVL	KRREEIEG	FDFA	GLVWL
5	orf125ng-1	LIGTVLAVMLPVT	EYENFLL	LIGSVFAPMA	AVLIADFFVL	KRREEIEG	FDFA	GLVWL
		310	320	330	340	350	360	
		360	370	380	390	400		
10	orf125-1.pep	FILYRFLSSG	WESSIGL	TAPVMSA	VAIATVSV	RLLFFKKT	QSLQ	RNP
	orf125ng-1	FILYRFLSSG	WESSIGL	TAPVMSA	VAIATVSV	RLLFFKKT	QSLQ	RNP
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

20	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
25	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.AC
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

	1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCL	NHTMMQENG	SLIVWHGQDK
	101	PLSSEFVRHL	KRGXTDDEI	VRWRADDIAE	REPQLGGRFX	DGIYLPTEXQ
	151	LDGRQLXSAL	ADALDELNPV	CHWEHECVPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	AAGCGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
40	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACC
	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
45	451	CTCGACGGGC	GGCAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCGAGCA	CACCAGCACC	CTGCGCGGCA	TACGCGGCGA
	651	AGTGCGCGCG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGTC
50	701	TGCTCCATCC	GCGTTATCCG	CTCTACATCG	CCCCGAAAGA	AAACACGTC
	751	TTCGTCATCG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CCCCGCCAG
	801	CGTGCGTTCA	GGGTTGGAAC	TCTGTGCCG	ACTCTATGCC	ATCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCCACG
	901	CTCAACCAAC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCGTGAT
55	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCCGTAA
	1001	CCGCCGCCGC	CGCCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGACGCG
	1051	CCCGAACGCG	ATAAAGAAAG	CGGTTTGCGC	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAML
51 PAEEAVEATP EVVRLGRQSI PLWRGIRCL NTHTMMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
201 WNQSPEHTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRRQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N.*

meningitidis:

```

15  orf126.pep      10      20      30      40      50      60
    orf126a      MTRIAILGGG LSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAXTVEATP
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf126a      MTRIAILGGG LSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAEAVEATP
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    10      20      30      40      50      60

20  orf126.pep      70      80      90      100     110     120
    orf126a      EVVRLGRQSIPLWRGIRCLNTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf126a      EVVRLGRQXIPLWRGIRCHLKTAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    70      80      90      100     110     120

25  orf126.pep      130     140     150     160     170     180
    orf126a      VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf126a      VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    130     140     150     160     170     180

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
51 ACTGCAGCTT GCAGAACAAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101 GCCGCCGGGG CGAACACGCC GCCGCCTATG TTGCCGCCGC CATGCTCGCG
35 151 CCTGCGGCGG AAGCGGTGCA AGCCACGCCT GAAGTGGTCA GGCTGGGCGAG
201 CGAGANCATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACNAAATC GTCGGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACC GC
40 401 AACTCGGCGG ACGTTTTC ACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
501 GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
601 TGGAACCAAT CCCCCANNA NACCAGCACC CTGCGCGGCA TACGCGGCGA
45 651 AGTGCGCGCG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTC
751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCGTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
50 901 CTCAATCACC ACAACCCGCA AATCCGTAC AACCGCGCCC GACGCCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCCGCGCG CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGANGCG
1051 CCCGAACGCG ATGAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA
1101 A

```

This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAML
51 PAEEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
60 201 WNQSPXXTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV

```

-448-

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
 301 LNHHNPEIRY NRARRLIEIN GLFRHGFMIS PAVTAAAVRL AVALFDGKXA
 351 PERDEESGLA YIRRQD*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5		10	20	30	40	50	60
	orf126a.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
	orf126-1	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
10		10	20	30	40	50	60
	orf126a.pep	70	80	90	100	110	120
	orf126-1	EVVRLGRQXIPLWRGIRCHLKTTPAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDDXI					
15		70	80	90	100	110	120
	orf126a.pep	130	140	150	160	170	180
	orf126-1	VRWRADDIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE					
20		130	140	150	160	170	180
	orf126a.pep	190	200	210	220	230	240
	orf126-1	DLQAQYDWLIDCRGYGAKTAWNQSPXXTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP					
25		190	200	210	220	230	240
	orf126a.pep	250	260	270	280	290	300
	orf126-1	LYIAPKENXVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIATGLRPT					
30		250	260	270	280	290	300
	orf126a.pep	310	320	330	340	350	360
	orf126-1	LNHHNPEIRYNRARRLIEINGLFRHGFMISPAVTAAAVRLAVALFDGKXAPERDEESGLA					
35		310	320	330	340	350	360
	orf126a.pep	310	320	330	340	350	360
	orf126-1	LNHHNPEIRYNRARRLIEINGLFRHGFMISPAVTAAAVRLAVALFDGKXAPERDEESGLA					
40		310	320	330	340	350	360
	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					
45							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEAVEATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
55	orf126ng	EVIRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
	orf126.pep	VRWRADDIAEREPLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE	180
	orf126ng	VRWRADEIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPO	180

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAML

-449-

51 PAEEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPO DLQAQYDWVI DCRGYAKTA
 201 WNQSPEHTST LRGIERGEVAR VYTPETITLNR PVRLHPRYP LYIAPKENHV
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSMPT STPPSAKPTS SKWRPGLRPT
 301 LNHNHPEIRY SRERRLIEIN GLFRHGFMIS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC
 10 51 ATTGCAGCTT GCAGAACAAG GTTATCAGAT TGAACCTTTC GACAAGGGCA
 101 CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG
 151 CCTGCGGCGG AAGCGGTCGA GGCAACGCCC GAAGTCATCA GGCTGGGCGAG
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGCTCTG AACACGCTCA
 251 CGATGATGCA GGAACACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
 15 301 CCATTATCCA GCGAGTTCTG CCGCCATCTC AAACGCGGCG GCGTAGCGGA
 351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC
 401 AACTCGGCGG ACCTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
 451 CTTGACGGGC GGCAAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
 501 GAACGTCCTT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG
 20 551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAACCGCGC
 601 TGGAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA
 651 AGTGGCGGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCGTGGCGCC
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC
 751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG
 25 801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG
 851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG
 901 CTCAACCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT
 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTCC CCGCCGTAA
 1001 CCGCCGCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG
 30 1051 CCCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA
 35 51 PAEEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPO DLQAQYDWVI DCRGYAKTA
 201 WNQSPEHTST LRGIERGEVAR VYTPETITLNR PVRLHPRYP LYIAPKENHV
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
 301 LNHNHPEIRY SRERRLIEIN GLFRHGFMIS PAVTAAAVRL AVALFDGKDA
 40 351 PERDEESGLA YIGRQD*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

		10	20	30	40	50	60
45	orf126-1.pep	MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
	orf126ng-1	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEAVEATP					
		10	20	30	40	50	60
50	orf126-1.pep	EVVRLGRQSIPLWRGIRCLNTHNTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
	orf126ng-1	EVIRLGRQSIPLWRGIRCLNLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
		70	80	90	100	110	120
55	orf126-1.pep	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE					
	orf126ng-1	VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPO					
		130	140	150	160	170	180
60	orf126-1.pep	GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPETITLNRPVRLHPRYP					
	orf126ng-1	DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPETITLNRPVRLHPRYP					
65		190	200	210	220	230	240

-450-

```

      250      260      270      280      290      300
orf126-1.pep  LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAIHPAFGEADILEIATGLRPT
5  orf126ng-1  LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPT
      250      260      270      280      290      300

      310      320      330      340      350      360
orf126-1.pep  LNHHNPEIRYNRRRLIEINGLFRHGFMI SPAVTAAARLAVALFDGKDAPERDKESGLA
10 orf126ng-1  LNHHNPEIRYSRERRLIEINGLFRHGFMI SPAVTAAVRLAVALFDGKDAPERDEESGLA
      310      320      330      340      350      360

orf126-1.pep  YIRRQDX
15 orf126ng-1  YIGRQDX

```

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

```

gi|2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
Length = 327
20 Score = 169 bits (423), Expect = 3e-41
   Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)

Query: 3   RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXXX 62
Sbjct: 2   RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEFPV 60

Query: 63  IRLGRQSIPLWRGIRCLNLTLMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
      + LGR +   W           +   G+L+V   G+D   F R   G   DE+
30 Sbjct: 61  LTLGRLAADWWEAA-----LPGHVHRRGTLVVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113

Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
      IA EP L GRF  ++   E LD RQ L+ALA L++   +
35 Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165

Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYPLY 242
      +D V+DC G           LRG+RGE+  V T E++L+RPVRLHPR+P+Y
40 Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTEVSLSRPVRLHPRHPIY 218

Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
      I P++ + F++GAT IES+   P + RS +ELL+A YA+HPAFGEA + E AG+RP
45 Sbjct: 219 IVPRDKNRFMVGATMIESDDGGPITARSLMELLNAAYAMHPAFGEARVTETGAGVRPAYP 278

Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMI SP 331
      + P R ++E R + +NGL+RHGF+++P
Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305

```

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

```

50 819>:
      1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
      51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
     101  TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
     151  CATTTTATGG AAAAGTTTAA TCTGCAGAAAT GGGAGGTTTA AACAAACATC
     201  TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC
     251  GTTTGAATGG AATCG+CGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
     301  AAGGCGGTAG CCATAGATAA AGATAAAAT CCTTTTATTA TTAAGATGAA
     351  TGAATATCTA GTAACCTTAA _TTTGAAGA AGTCCGCCAG TTCGTGTAGT
     401  GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
     451  GTAG
60

```


-451-

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
101 KAVAIKDKRN PFIIMNENL VTFICKKSAS SCSGLDYFK GNDKDKLLK
151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGC GGGAGGTTA ACAAACATC
151 CATTTTATGG AAAAGTTTTA TCTGCAGAA GGGAGGTTA ACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
151 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIKDKKNP FIIKMENLV TFICKKSASS CSDGLDYFKG NDKDKLLK*

```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N. meningitidis*:

```

25 orf127.pep 10 20 30 40 50 60
MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN HFMEKFYLN
|||||:|||||
orf127a MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN HFMEKFYLN
10 20 30 40 50 60

30 orf127.pep 70 80 90 100 110 120
GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR XALDSKFMLK AVAIKDKKNP FIIKMENLV
|||||:|||||
orf127a GRFKQTSTKW PSLPIKEAEG FCIRLNGI-ARGALDSKFMLK AVAIKDKKNP FIIKMENLV
70 80 90 100 110

35 orf127.pep 130 140 150
VTFICKKSASS CSDGLDYFKG NDKDKLLKX
|||||
orf127a VTFICKKSASS CSDGLDYFKG NDKDKLLKX
120 130 140 150

```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT ACAGTGC GGGAGGTTA ACAAACATC
45 151 CATTTTATGG AAAAGTTTTA TCTGCAGAA GGGAGGTTA ACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
50 401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIKDKKNP FIIKMENLV TFICKKSASS CSDGLDYFKG NDKDKLLK*

```

-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

```

      10      20      30      40      50      60
orfl27a.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLEN AHFMEKFYLQN
5 orfl27-1   MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl27a.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
10 orfl27-1   GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90     100     110     120

      130     140     150
orfl27a.pep TFICKKSASSCS DGLDYFKGNDKDKLLKX
15 orfl27-1   TFICKKSASSCS DGLDYFKGNDKDKLLKX
      130     140     150

```

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

```

orfl27.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN 60
25 orfl27ng MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFL ENAHFMEKFYLQN 60

orfl27.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARXALDSKFMLKAVAIDKDKNPFI IKMNENL 120
30 orfl27ng GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFI IKMNENL 119

orfl27.pep VTFICKKSASSCS DGLDYFKGNDKDKLLK 150
orfl27ng VTFICKKSASSCS DRLDYFKGNDKDKLLK 149

```

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

```

35 1 ATGACTGATA ATCGGGGGTT TACACTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GACTTGTCT TAATTGTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGC GGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTT TCTGCAGAAAT GGGAGATTTA AACAAACATC
201 TACCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
40 251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 826>:

```

45 1 MTDNRGFTLV ELISVVLILSVLALIVYPSY RNYVEKAKIN AVRAAFL ENA
51 HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIIKMNNENLV TFICKKSASS CSDRLDYFKG NDKDKLLK*

```

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

```

      10      20      30      40      50      60
50 orfl27-1.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN
orfl27ng-1 MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN
      10      20      30      40      50      60

      70      80      90     100     110     120
55 orfl27-1.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
orfl27ng-1 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90     100     110     120
60

```

-453-

```

              130      140      150
orf127-1.pep  TFICKKSASSCS DGLDYFKGNDKDCKLLKX
              |||
orf127ng-1    TFICKKSASSCS DGLDYFKGNDKDCKLLKX
              130      140      150

```

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

```

1  ..GTGTCGCTGG CTTCCGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
51  CAACCAAATG CCGAAAACCG GGGAGCTAT CTGCGGTTTT CTGTCCAAT
101 ATTTATCTGG GGTTCAGCA GGGGTATTTC GATTTGAGTG CCGACGAGAA
151 CCCCCTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
201 TGTATCCCTT TTTGCTGATA TTTTGCTGCA AAAAAACCAA ATCGCTACGG
251 GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
301 GTTTTTGCCA AGCGGGTTT ATACCGACAT CCTCAACCAA CCAATACTT
351 ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTTGGCAGG TTCGCTGCTG
401 GCGGTTTACG GGCAAACGCA AAACGGCAGA CCGCAAACAG CAAATGGAAA
451 ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
501 TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
551 CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
601 TCCGACCCGC ATCCTGTCCG CAAGCCCCAT CGTATTTGTC GGCAAAATCT
651 CTTATTCCCT ATACCTGTAC CATTGGATT TTATTGCTTT CGCTCCGCTC
701 ATTAGAGGCG GGAAACAGCT CGGACTGCCT GCCG..

```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```

1  ..VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGFAQ GYFDLSADEN
51  PVLHIWSLAV EEQYLLLYPL LLIFCCKKTK SLRVLRNISI ILFLILTASS
101 FLPSGFYTDI LNQPNTYYLS TLRFPPELLAG SLLAVYGQTQ NGRRQTANGK
151 RQLLSSLCFG ALLACLVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
201 PTRILSASPI VVFGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

```

1  ATGCAAGCTG TCCGATACAG ACCGGAATTT GACGGATTGC GGGCCGTCGC
35  CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
151 GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
201 TTATACCCGC AGGATTAAGC GGATTTATCC TGCCTTTATT GCGGCCGTGT
40  CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTC AAC
301 CAAATGCGGA AAACCGTGGA GCTTTCTGCG GTTTTCTTGT CCAATATTTA
351 TCTGGGGTTT CAGCAGGGGT ATTTCGATTG GAGTGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
451 CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT
501 GCGTAACATC AGCATCATCC TGTTTTTGAT TTTGACTGCC TCATCGTTTT
45  TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCTA TACTTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTTCG TGCTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC
701 AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCTCG CCTGTTCTGT
751 ATTGACAAAC ACAATCCGTT TATCCGGGGA ATGACCCTGC TCCTTCCCTG
50  CCTGCTGACG CCACTGCTTA TCCGGAGTAT GCAATACGGG AACTTCCGA
801 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTT GATTTTATTG GCTTTCGCCC ATTACATTAC
951 AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA
1001 CGGCCGGATT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA
55  AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTCT ATCTCGCCCC
1101 GTCCCTGATA CTTGTCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCCGTT CCCGGCGCGC CCCTTGCTGC GGAAAATCAT

```

-454-

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185

TTTCCGGAAA CCGTCCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGAG
GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAA GCCAAAATCC
TGTCCTCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
AACCCGTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT
TTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCGA
GATTTGAAGC GCAATCCTTC CTAATACCCG GGTCCCAGC CCGATTGAGG
GAAACCGTCA AAAGGATAGC CGCCGTCAAA CCCGTCTATG TTTTTCGAAA
CAACACATCA ATCAGCCGTT CGCCCCTGAG GGAGGAAAAA TTGAAAAGAT
TTGCCGCAAA CCAATATCTC CGCCCCATTC AGGCTATGGG CGACATCGGC
AAGAGCAATC AGGCGGTCTT TGATTGATT AAAGATATTC CCAATGTGCA
TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
TATATGGGCG GGAATTCCA CAAACACGAA CGCCTGCTTA AATCTTCCCA
CGGCGGCGCA TTGCACTAG

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT
51 GIILSEIQNG SFSFRDFYTR RIKRIYPAPI AAVSLASVIA SQIFLYEDFN
101 QMRKTVELSA VFLSNIYLG FQQGYFDLSAD ENPVLHIWSL AVEEQYLLY
151 PLLLIFCCKK TKSLRVLNRI SIILFLILTA SSFLPSGFYT DILNQNTYY
201 LSTLRFPELL AGSLLAVYGQ TQNGRRQTAN GKROLSSLC FGALLACLFV
251 IDKHNPFIPIG MTLPLCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGEFSL SYYLIEQPLR
351 KRKMTFFKAF FCLYLAPSLI LVGYNLYARG ILKQEHRLPL PGAPLAAENH
401 FPETVLTLDG SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFARFR
501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
551 KSNQAVFDLI KDIPNVHVD AOKYLPKNTV EIYGRYLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKSSHGGA LQ*

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723)

ORF128 and HI0392 show 52% aa identity in 180aa overlap:

35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635
8640
8645
8650
8655
8660
8665
8670
8675
8680
8685
8690
8695
8700
8705
8710
8715
8720
8725
8730
8735
8740
8745
8750
8755
8760
8765
8770
8775
8780
8785
8790
8795
8800
8805
8810
8815
8820
8825
8830
8835
8840
8845
8850
8855
8860
8865
8870
8875
8880
8885
8890
8895
8900
8905
8910
8915
8920
8925
8930
8935
8940
8945
8950
8955
8960
8965
8970
8975
8980
8985
8990
8995
9000
9005
9010
9015
9020
9025
9030
9035
9040
9045
9050
9055
9060
9065
9070
9075
9080
9085
9090
9095
9100
9105
9110
9115
9120
9125
9130
9135
9140
9145
9150
9155
9160
9165
9170
9175
9180
9185
9190
9195
9200
9205
9210
9215
9220
9225
9230
9235
9240
9245
9250
9255
9260
9265
9270
9275
9280
9285
9290
9295
9300
9305
9310
9315
9320
9325
9330
9335
9340

-455-

5	orf128a	ILFLILTATSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGQTONGRRQTANGK	180	190	200	210	220	230
	orf128.pep	RQLLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	160	170	180	190	200	210
10	orf128a		240	250	260	270	280	290
	orf128.pep	RQLLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	220	230	240			
15	orf128a	VFVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA	300	310	320	330	340	350
	orf128.pep							
	orf128a	VFVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSVAALTAGFSLLSYYLIEQPLRKR	360	370	380	390	400	410
	orf128.pep							

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
51	CGTGCTATCC	GTCTGATATT	TCCACCTGAA	TAACCGCTGG	CTGCCCCGGAG
101	GATTCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
151	GGCATCATT	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTT	TCCGGGATTT
201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
301	CAAATGCGGA	AAACCGTGGA	GCTTCTGCG	GTTTCTTGT	CCAATATTTA
351	TCTGGGGTTT	CAGCAGGGGT	ATTTCGATT	GAGTGCCGAC	GAGAACCCCG
401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
501	GCGTAACATC	AGCATCATCC	TATTTCTGAT	TTGACTGCC	ACATCGTTTT
551	TGCCAAGCGG	GTCTTATACC	GATATTCTCA	ACCAACCCAA	TACTTATTAC
601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTCGC	TGCTGGCGGT
651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAAACGGC
701	AGTTGCTTTC	ATCACTCTGC	TTCCGGCGCAT	TGCTTGCGTG	CCTGTTCTGT
751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCCTGC	TCCTTCCCTG
801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
901	TCCCTATACC	TGTACCATTC	GATTTTATT	GCTTTCGCC	ATTACATTAC
951	AGGCGACAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
1001	CGGCCGGATT	TTCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATT	TTCTGCCTCT	ATCTCGCCCC
1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGA AAC
1151	AGGAACACCT	CCGCCCGTTG	CCCGGCGCGC	CCCTTGCTGC	GGAAAATCAT
1201	TTTCCGGAAA	CCGTCCTGAC	CCTCGGCGAC	TGCGACGCCG	GACACCTGCG
1251	GGGGTTTCTG	GATTATGTCTG	GCAGCCGGGA	AGGGTGGAAA	GCCAAAATCC
1301	TGTCCCTCGA	TTCGGAGTGT	TTGGTTTGGG	TAGATGAGAA	GCTGGCAGAC
1351	AACCCGTTAT	GTCGAAAATA	CCGGGATGAA	GTTGAAAAAG	CCGAAGCCGT
1401	TTTCATTGCC	CAATTCTATG	ATTTGAGGAT	GGGCGGCCAG	CCCGTGCCGA
1451	GATTTGAAGC	GCAATCCTTC	CTAATACCCG	GGTCCCAGC	CCGATTTCAGG
1501	GAAACCGTCA	AAAGGATAGC	CGCCGTCAAA	CCCGTCTATG	TTTTTGCAAA
1551	CAACACATCA	ATCAGCCGTT	CGCCCCTGAG	GGAGGAAAAA	TTGAAAAGAT
1601	TTGCCGCAAA	CCAATATCTC	CGCCCCATTC	AGGCTATGGG	CGACATCGGC
1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATTC	CCAATGTGCA
1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
1751	GCCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTTCTTAT
1801	TATATGGGGC	GGGAATTTCA	CAAACACGAA	CGCCTGCTTA	AATCTTCTCG
1851	CGACGGCGCA	TTGCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

1	MQAVRYRPEI	DGLRAVAVLS	VMIFHLNNRW	LPGGFLGVDI	FFVISGFLIT
51	GIILSEIQNG	SFSFRDFYTR	RIKRIYPAFI	AAVSLASVIA	SQIFLYEDFN
101	QMRKTVELSA	VFLSNIIYLG	QQGYFDLSAD	ENPVLHIWSL	AVEEQYLLLY
151	PLLLIFCCKK	TKSLRVLRLNI	SIILFLILTA	TSFLPSGFYT	DILNQPNYY
201	LSTLRFPELL	AGSLLAVYGQ	TONGRRQTAN	GKROLSSSLC	FGALLACLFV
251	IDKHNPFIPG	MTLLLPCLLT	ALLIRSMQYG	TLPTIRLSAS	PIVFGKISY
301	SLYLYHWIFI	FAFAHYITGDK	QLGLPAVSAV	AALTAGFSL	SYYLIEQPLR
351	KRKMTFKKAF	FCLYLAPSLI	LVGYNLYARG	ILKQHLRPL	PGAPLAAENH
401	FPETVLTGLD	SHAGHLRGFL	DYVGSREGWK	AKILSLDSEC	LWVWDEKLAD
451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFPARFR

501 ETVKRIA AVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPNVHWVD AQKYLPKNTV EIYGRYLYGD QDHLTYFGSY
 601 YMGREFHKHE RLLKSSRDGA LQ*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128a.pep	QQGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
	orf128-1	TSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
20	orf128a.pep	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
	orf128-1	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLDADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLDADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFILPGFPARFRET VKRIA AVKPVYVFANNTSISRSPLREEK LKRFAANQYL
	orf128-1	PVPRFEAQSFILPGFPARFRET VKRIA AVKPVYVFANNTSISRSPLREEK LKRFAANQYL
40	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDI PNHVHWDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDI PNHVHWDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGALQX

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N.gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAV	30
55	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
	orf128.pep	LSNIYLG FQQGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISI	90
	orf128ng	LSNIYLG FRLGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCYKTKTKSLRVLNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGK	232
65	orf128.pep	RQLLSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
	orf128ng	RQLLSLLCFGALLVCLFVIDKHDPFI PGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292

-457-

orf128.pep VFVVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA 244
 orf128ng VFVVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR 352

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

```

1  ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGC
51  CGTGTATCC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCCTGGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC
151 AACATCATT TTTCTGAAAT ACAGAACGGT TCTTTTTCTT TCCGGGATTT
10  201 TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTATT GCGGCCGTGT
251 CCTCGGCTTC GGTGATTGCT TCTCAAATCT TCCTTTACGA AGATTTC AAC
301 CAAATGAGGA AAACCATAGA GCTTTCTACG GTTTTTTGT CCAATATTTA
351 TTTGGGGTTC CGATTGGGGT ATTTCGATTT GAGTGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT
15  451 CCTCTTTTGC TGATATTCTG TTACAAAAAA ACCAAATCAC TACGGGTGCT
501 GCGTAATATC AGCATCATCC TGTTTCTGAT TTTGACCGCA TCATCGTTTT
551 TGCCGGCCGG GTTTTATACC GACATCCTCA ACCAACCCaa TACTTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTTCG TGTTGCGGGT
15  651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAT GGAACCGGC
20  701 AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTTGTCTG CCTGTTCTGT
751 ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCCTGC TCCTTCCCTG
801 CCTGCTGACG GCGTGCTTA TCCGGAGTAT GCAATACGGG AACTTCCGA
851 CCGCATCCT GTCCGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTG GATTTTATT GCCTTCGCCC ATTACATTAC
25  951 AGGCGACAAA CAGCTCGGAC TGCTGCGGT ATCGGCGGTT GCCGCGTTGA
1001 CGGCGGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTAGA
1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTTT ATCTCGCCCC
1101 GTCCCTGATG CTTGTCGGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCGCTG CCCGGCACGC CCGTTGCTGC GGAAAATAAT
30  1201 TTTCCGAAA CCGTCTTGAC CCTCGCGGAC TCGCACGCCG GACACCTGCG
1251 GGGGTTTCTG GATTATGTCT GCGGCAGGGA AGGGTGGAAA GCTAAAATCC
1301 TGTCCCTCGA TTCGGAGTGT TTGTTTGGG TGATGAGAA GCTGGCAGAC
1351 AATCCGTTGT GCCGAAAATA CCGGGATGAA GTTGA AAAAG CCGAAGCTGT
1401 TTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
35  1451 GATTTGAAGC GCAATCCTTC CTGATACCCG GGTCAAAGC CCGATTACAG
1501 GAAACCGTCA AGAGGATAGC CGCCGTCAA CCTGTATATG TTTTGTGAAA
1551 CAATACATCA ATCAGCCGTT CTCCCTTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCTATAAA CCAATACCTC CGGCCTATT GGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGGCGGTCTT TGATTGGTT AAAGATATTC CCAATGTGCA
40  1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG
1751 GACGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGGC GGAATTTCA CAAACACGAA CGCCTGCTCA AGCATCCCCG
1851 AGGCGGCGCA TTGCAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 834>:

```

45  1  MQAVRYRPEI DGLRAVAVLS VIIFHLNNRW LPGGFLGVDI FFVISGFLIT
51  NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTIELST VFLSNIYLG FRLGYFDLSAD ENPVLHIWSL AVEEQYLLY
151 PLLLIFCYKK TKSRLVLRNI SIILFLILTA SSFLPAGFYT DILNQPNYY
50  201 LSTLRFPELL VGSLLAVYQ TQNGRRQ TEN GKRQLSLLC FGALLVCLFV
251 IDKHDPFIPG ITLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR
351 KRKMTFFKAF FCLYLAPSLM LVGYNLYSRG ILKQEHRLPL PGT PVAENN
401 FPETVLTLDG SHAGHLRGFL DYVGGREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFKARF
55  501 ETVKRIA AVK PVIYFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDIG
551 KSNQAVFDLV KDIPNVHWDV AQKYL PKNTV EIHGRYLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKHSRGA LQ*

```

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

```

60  orf128-1.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
orf128ng MQAVRYRPEIDGLRAVAVLSVIIFHLNNRWLPGGFLGVDIFFVISGFLITNIILSEIQNG
orf128-1.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
65  orf128ng SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG

```

5	orf128-1.pep	QQGYFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA :
	orf128ng	RLGYFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCYKTKSLRVLNRNISIILFLILTA
10	orf128-1.pep	SSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC :
	orf128ng	SSFLPAGFYTDILNQPNYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGKRQLLSSLC :
15	orf128-1.pep	FGALLACLFVIDKHNPFIPIGNTLLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY :
	orf128ng	FGALLVCLFVIDKHDPFIPIGITLLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY :
20	orf128-1.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF :
	orf128ng	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF :
25	orf128-1.pep	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL :
	orf128ng	FCLYLAPSLMLVGYNLYSRGILKQEHRLPLPGTPVAAENHFPETVLTGLGDSHAGHLRGFL :
30	orf128-1.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ :
	orf128ng	DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ :
35	orf128-1.pep	PVPRFEAQSFILIPGFARFRETVKRIAARKPVYVFANNTSISRSPREEKLKRFAANQYL :
	orf128ng	PVPRFEAQSFILIPGFARFRETVKRIAARKPVYVFANNTSISRSPREEKLKRFAINQYL :
40	orf128-1.pep	RPIQAMGDIGKSNQAVFDLIKDI PNHVWVDAQKYLKNTVEIYGRYLYGDQDHLTYFGSY :
	orf128ng	RPIRAMGDIGKSNQAVFDLVKDI PNHVWVDAQKYLKNTVEIHGRYLYGDQDHLTYFGSY :
45	orf128-1.pep	YMGREFHKHERLLKSSHGGALQX :
	orf128ng	YMGREFHKHERLLKSHRGALQX :

610 620

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

40	sp P43993 Y392_HAEIN	HYPOTHETICAL PROTEIN HI0392 >gi 1074385 pir B64007
	hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)	
45	>gi 1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus influenzae]	Length = 245
	Score = 239 bits (604), Expect = 3e-62	Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)
50	Query: 38	VDIFFVISGFLITNIIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
	Sbjct: 1	MDIFFVISGFLITIIITEIQNSFSLKQFYTRRIKRIYPAFITVMALVSFIASAFIYIN 60
55	Query: 98	DFNQMRTIELSTVFLSNIIYLGFRGLGYFDLSADENPVLHIWISLAVEEQXXXXXXXXXIFC 157
	Sbjct: 61	DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWISLAVE Q I 120
60	Query: 158	YKKTSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPELLVGSLLAV 217
	Sbjct: 121	YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+ 180
65	Query: 218	YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDPFIPIGIT 262
	Sbjct: 181	Y N + Q +L++L L CLF+++ + FIPGIT 224

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT
5  101  TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGCGC
151  GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201  CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTGTTGGCG TATGTGTGGT
251  TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10  1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTTC GTTTTGACAT TATTACGAA TACCGCTGGA TGTTTCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
15  101  CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTGGAAAAAA
151  GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
201  GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251  TTTGGGCGTA TGTGTGGTTT CCGTTTTCG TCCATCCTTC AGACGGCATT
301  TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
20  351  GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401  AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451  GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501  GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551  CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGAGTTG
25  601  GCGTATGTTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651  GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701  GGATATTCCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

1  MDRFRDIIYE YRWMFLYAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
30  51  AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVHPSDGI
101  LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151  ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
201  AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

40  orf129.pep      10      20      30      40      50
      I IYEYRWMFLYALTTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      |||
orf129a  MDRFRDIIYEYRWMFLYALTTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      10      20      30      40      50      60

45  orf129.pep      60      70      80
      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFFV
      |||
orf129a  ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFFVHPSDGI LVSGEAAIALRRGYGPLIAG
      70      80      90      100     110     120

50  orf129a      SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS
      130     140     150     160     170     180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTTC GTTTTGACAT TATTACGAA TACCGCTGGA TGTTTCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

```

-460-

5
10
15
20
25
30
35
40
45
50
55
60
65
70

```

101 CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301 TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451 GCGCGTTCTT TGGGGCTGAC CTATCCGCG GCGATGCGCT ATGTGATTCT
501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG
601 GCGTATGTTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTAGGCT
701 GGATATTCCT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 840>:

15
20
25
30
35
40
45
50
55
60
65
70

```

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
201 AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

25
30
35
40
45
50
55
60
65
70

```

orf129a.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
orf129-1      MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW

orf129a.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG
orf129-1      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG

orf129a.pep SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS
orf129-1      SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS

orf129a.pep EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE
orf129-1      EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE

orf129a.pep KRYNPQHRX
orf129-1      KRYNPQHRX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120

```

orf129.pep      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW 54
orf129ng        MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 60

orf129.pep      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFV 88
orf129ng        ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVILHTAFLGNAMQSRVPDKGRWIAG 120

```

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55
60
65
70
75
80
85
90
95
100
105
110
115
120

```

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVILHTAF
101 LGNAMQSR RVPDKGRWIAG SLELNCQPRG RKTRGEFPPG ESNLGTPEPN
151 PLSMGQRRFP GCENWYPPQN FIKK*

```

Further work revealed the following gonococcal sequence <SEQ ID 843>:

```

1 ATGGATTTTc gtTTTGACAT TATTTAcgaA TACCGCTGGA TGTTTCTTTA

```

-461-

5
10
15
20
25
30
35
40
45
50
55
60
65
70

```

51  CGGCGCACTG  Acgaccttgg  ggctgacggt  cgtggcgacg  gCGGGCGGTT
101 CGGtatttggG  TCTGTTGTTG  GCGTTGGCGC  GCCTGATTCA  CTTGGAAAAA
151 GCCGGTGCGC  CGATGCGCGT  GCTGGCGTGG  GCGTTGCGTA  AGGTTTCGCT
201 GCTGTACGTT  ACCCTGTTCC  GGGGTACGCC  GCTGTTTGTG  CAGATTGTGA
251 TTTGGGCGTA  TGTGTGGTTT  CCGTTTTTCG  TCCATCCTTC  AGACGGCATT
301 TTGGTCAGCG  GCGAGGCGGC  AATCGCGCTG  CGTCGCGGAT  ACGGGCCGCT
351 GATTGCCGGT  TCTTTGGCAC  TGATCGCCAA  CTCGGGGGCG  TATATCTGTG
401 AGATTTTCCG  CGCGGGCATC  CAGTCTATAG  ACAAAGGACA  GATGGAGGCG
451 GCGTGTCTT  TGGGACTGAC  CTATCCGCG  GCGATGCGCT  ATGTGATTCT
501 GCCGCAGGCA  TTGCGCCGTA  TGCTGCCGCC  TTTGGCGAGC  GAGTTCATCA
551 CGCTCTTGAA  AGACAGCTCG  CTGCTGTCGG  TCATTGCTGT  GCGGGAGTTG
601 GCGTATGTT  AGAATACGAT  TACGGGCGCG  TATTCGGTTT  ATGAAGAACC
651 GCGTTACACC  GCCGCCCTGA  TTTATCTGTT  GATGACGACT  TTCTTAGGCT
701 GGATATTCCT  GCGTTTGGAA  AAACGTTACA  ATCCGCAACA  CCGCTGA

```

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

20
25

```

1  MDRFRDIIYE  YRWMFLYGAL  TTLGLTVVAT  AGGSVLGLLL  ALARLIHLEK
51  AGAPMRVLAW  ALRKVSLLYV  TLFRTGPLEV  QIWIWAYVWF  PFFVHPSDGI
101 LVSGEAAIAL  RRGYGPLIAG  SLALIANSQA  YICEIFRAGI  QSIDKGQMEA
151 ARSLGLTYPQ  AMRYVILPQA  LRRMLPPLAS  EFITLLKDSS  LLSVIAVAEL
201 AYVQNTITGR  YSVYEEPLYT  VALIYLLMTT  FLGWIFLRLE  KRYNPQHR*

```

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

25
30
35
40

```

orf129-1.pep  MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
orf129ng-1    MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW

orf129-1.pep  ALRKVSLLYVTLFRGTPLFVQIWIWAYVWFFVHPSDGILVSGEAAIALRRGYGPLIAG
orf129ng-1    ALRKVSLLYVTLFRGTPLFVQIWIWAYVWFFVHPSDGILVSGEAAIALRRGYGPLIAG

orf129-1.pep  SLALIANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS
orf129ng-1    SLALIANSQAYICEIFRAGIQSIDKGQMEAAACSLGLTYPQAMRYVILPQALRRMLPPLAS

orf129-1.pep  EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
orf129ng-1    EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLE

orf129-1.pep  KRYNPQHRX
orf129ng-1    KRYNPQHRX

```

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

45
50
55

```

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
[Archaeoglobus fulgidus]Length = 224
Score = 132 bits (329), Expect = 2e-30
Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

Query: 65  VSLLYVTFLFRGTPLFVQIWIWAYVWFFVHPSDGILVSGEAAIALRRGYGPLIAGSLAL 124
      +S  YV + RGTPL VQI+I      +F  P+ GI + E A      G +AL
Sbjct: 58  ISTAYVEVIRGTPLLQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99

Query: 125  IANSQAYICEIFRAGIQSIDKGQMEAAACSLGLTYPQAMRYVILPQALRRMLPPLASEFIT 184
      SGAYI EI RAGI+SI  GQMEAA SLG+TY QAMRYVI PQA R +LP L +EFI
Sbjct: 100  SICSGAYIAEIVRAGIESIPIGQMEAAARSLGMYTYLQAMRYVIFPQAFRNILPALGNEFIA 159

Query: 185  LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLEKR 242
      LLKDSSLLSVI++ EL V I      P  AL YL+MT L + +K+
Sbjct: 160  LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLLMTIPLSRLVAYSQKK 217

```

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1   ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
101 GGCTGCCCCG GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151 CTGCTCGCCA AGCTGCGTGA gCTTACCACAT CACGAACCTCT TACGTAAACA
201 cTACGTCCGC ACTTATTACy TGCTCCAACCT CTTTGCCGCC GCAGgcTagT
251 TTGTGGACAG GCGCGCGGwa ATTACAAAAC CTGCCCCGcyT CCGCGCCCTT
301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
10 351 TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACCTCGA CTACCCCAA
401 CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
451 TTTCTTGrTG AACGTGAACC CGrTATTTTT CATTACCGTT CCTGCGATTTC
501 TGACCGCCGC CGTATTCGTA CTGTATCTTT TCrCGTTTAT ACCGATATTT
551 CGGGCGAATG CGTTTACAGA CGATCCGGAr TAr

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1   ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPASAPL
101 HLITLGGMMG GVMMVWLTAG LWHSGFTKLD YPKLCRIAVP ILFAAAVSRA
151 FLXNVNPXFF ITVPAILTAA VEVLYLFXFI PIFRANAFTD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1   ATGCGGCCGT TTTTCGTCGG CGCGCGGGTG CTTGCCATAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101 TGGAACCTAT GCTGCCGGCG GCATACGGCG GTTTTTTGAC TCGGCTTTTG
151 TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCTG CGACTTTGAT
25 201 GCGGCGATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
251 CTGCTCGGTT TTTCTGTCGC GCCTATTGGC TGGTGTGCTG GCTGTTCTGC
301 GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCTCGCTAAT
351 GTTACTTGCC GCGTTCACTG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
401 ATTTGAACCT GTTGCGCGCG CAAGTGCATC TAAATATGGC GCGCGTGATG
10 451 TTCGTATCCG TGCGCGTCAG TATTCTTTTG GCGCGGAAG CCCTGAAAGA
501 ATGCCGTCTG AAAGACCCTG TTTTATTATCC AAATATCGTT TATAAAAACA
551 TCGCCATTAC TTTCTGCTC TTGCACGCCG CCGCCGAAC TTGGCTGCCC
601 GCGCAAACCG CCGGTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
651 CAAGCTGCGT GAGCTTCACC ATCACGAAC CTTACGTAAC CACTACGTCC
35 701 GCACTTATTA CTGTCTCAA CTCTTTGCCG CCGCAGGCTA TTTGTGGACA
751 GCGCGGGCGA AATTACAAAA CCTGCCGCCG TCCGCGCCCC TGCACCTGAT
801 TACCCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCCG
851 GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCAA ACTCTGCCGC
901 ATTGCCGTCC CCATCCTTTT CGCCGCCGCC GTCTCGCGCG CTTTCTTGAT
40 951 GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCCG
1001 CCGTATTCGT ACTGTATCTT TTCACGTTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1   MRPFFVGAAV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
45 51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLILLFC
101  ARLIWLDRNT DNFALLMLLA AFTVFQYAYA VSGDLNLLRA QVHLNMAAVM
151  FVSVRVSILL GAEALKECRL KDPVFIPNIV YKNIAITFLL LHAAELWLPL
201  AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYYLLQ LFAGAGYLWT
50 251  GAAKLQNLPA SAPLHLITLG GMMGGVMMVW LTAGLWHS GF TKLDYPKLCR
301  IAVPILFAAA VSRAFLMNVN PIFFITVPAI LTAAVFVLYL FTFIPIFRAN
351  AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N. meningitidis*:

-463-

```

                                10      20      30
orf130.pep                      LKECRLKDPVFIPNIVYKNIAITFLLHAA
                                |||||
orf130a      LNLRLAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNVYKNIAITFLLHAA
                    140      150      160      170      180      190

                                40      50      60      70      80      90
orf130.pep      AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX
|||
orf130a      AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAX
                    200      210      220      230      240      250

                                100      110      120      130      140      150
orf130.pep      LQNLPAAPLHLITLGGMMGVMVWLTAGLWHSGF TKLDYPKLCRIAVPILFAAAVSRA
|||
orf130a      LQNLPAAPLHLITLGGMMGVMVWLTAGLWHSGF TKLDYPKLCRIAVPILFAAAVSRA
                    260      270      280      290      300      310

                                160      170      180      190
orf130.pep      FLXNVNPFITVPAILTAADFVLYLFXFIPIFRANAFTDDPEX
|||
orf130a      VLMNVNPIFFITVPAILTAADFVLYLLTFVPIFRANAFTDDPEX
                    320      330      340      350

```

The complete length ORF130a nucleotide sequence <SEQ ID 849> is:

```

25      1  ATGCGGCCGT  TTTTCGTCGG  CGCGGCGGTG  CTTGCCATAC  TCGGTGCGCT
      51  GGTGTTTTTC  ATCAACCCCG  GTGCCATCGT  CCTGCACCGC  CAAATTTTCT
    101  TGGAACTTAT  GCTGCCGGCG  GCATACGGCG  GTTTTTTGAC  TCGCGCTTTG
    151  TTGGACTGGA  CGGGTTTTTC  GGGTAACCTG  AAACCTGTCTG  CGACTTTGAT
    201  GCGCGCATTA  TTGCTCGCCG  CATCCGCTAT  ACTGCCCTTT  TCGCCGCAAA
    251  CTGCCTCGTT  TTTTCGTCGC  GCCTATTGGC  TGGTGTGTCT  GCTGTTCTGC
    301  GCGCGGCTGA  TTTGGCTAGA  CCGAAACACC  GACAACTTCG  CCCTGCTAAT
    351  GTTACTTGCC  GCGTTCACGT  TTTTTCAGAC  GGCATATGCC  GTCAGCGGCG
    401  ATTTGAACCT  GTTGCCTGCG  CAAGTGCATC  TAAATATGGC  GCGCGTGATG
    451  TTCGTATCCG  TGCGCTCAG  TATTCTTTTG  GGCGCGGAAG  CCCTGAAAGA
    501  ATGCCGTCTG  AAAGACCCAG  TATTCATCCC  CAATGTCTGC  TATAAAACA
    551  TCGCCATTAC  CTTCTGCTC  CTGCACGCCG  CCGCCGAACT  TTGGCTGCCT
    601  GCGCAAACCG  CCGGTTTTAC  CTCGCTCGCC  GTCGGCTTTA  TCCTGCTTGC
    651  CAAGCTGCGT  GAGCTTCACC  ATCACGAACT  CCTGCGCAAA  CACTACGTCC
    701  GCACTTATTA  CTTGCTCCAA  CTCTTTGCCG  CCGCAGGCTA  TTTGTGGACA
    751  GGCGGGCGGA  AATTACAAAA  CCTGCCCGCC  TCCGCGCCCC  TGCACCTGAT
    801  TACCCTCGGT  GGCATGATGG  GCAGCGTGAT  GATGGTGTGG  CTGACTGCGC
    851  GACTGTGGCA  CAGCGGCTTT  ACCAAGCTCG  ACTACCCGAA  ACTCGCCGCG
    901  ATCGCCGTCC  CCATCCTNTT  CGCCGCGGCC  GTTTCGCGCG  CTGTTTAAAT
    951  GAACGTAAAC  CCGATATTCT  TCATCACCGT  CCCCAGCAAT  CTGACCGCCG
   1001  CCGTGTTCTG  GCTTTACCTG  CTGACATTCG  TACCGATCTT  TCGGGCGAAC
   1051  GCGTTTACAG  ACGATCCGGA  ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 850>:

```

      1  MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
    51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLVLLFC
   101  ARLIWLDRNT DNFALLMLLA AFTVFQYAYA VSGDLNLLRA QVHLNMAAVM
   151  FVSVRVSILL GAEALKECRL KDPVFIPNVV YKNIAITFLL LHAAAEWLWP
   201  AQTAGFTSLA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAGAGYLWT
   251  GAAKLQNLPA SAPLHLITLG GMMGSVMVW LTAGLWHSGF TKLDYPKLCR
   301  IAVPILFAAA VSRVLMNVN PIFFITVPAL LTAADFVLYL LTFVPIFRAN
   351  AFTDDPE*

```

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

```

orf130a.pep      MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL LDWTGFSGNL
|||
orf130-1          MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL LDWTGFSGNL

orf130a.pep      KPVATLMAALL LAASAILPF SPQTASFFVA AYWLVLVLLFC ARLIWLDRNT DNFALLMLLA
|||
orf130-1          KPVATLMAALL LAASAILPF SPQTASFFVA AYWLVLVLLFC ARLIWLDRNT DNFALLMLLA

orf130a.pep      AFTVFQYAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNVV

```

	orf130-1	 AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAELKECRLKDPVFIPNIV
5	orf130a.pep	YKNIAITFLLHAAAELWLPAGTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLO
	orf130-1	 YKNIAITFLLHAAAELWLPAGTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLO
	orf130a.pep	LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGVMVWLTAGLWHSFGFTKLDYPKLCR
10	orf130-1	 LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGVMVWLTAGLWHSFGFTKLDYPKLCR
	orf130a.pep	IAPVILFAAAVSRAVLNMVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPE
15	orf130-1	 IAPVILFAAAVSRAFLNMVNPIFFITVPAILTAAVFVLYLTFPIPIFRANAFTDDPE

Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

20	orf130.pep	LKECRLKDPVFIPNIVYKNIAITFLLHAA	30
	orf130ng	LNLLRAQVHLNMAAVMFVSVRVSLLGTETLKECRLKDPVFIPNVIYKNIAIT-LLHAA	201
	orf130.pep	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX	90
25	orf130ng	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX	261
	orf130.pep	LQNLPAAPLHLITLGGMMGVMVWLTAGLWHSFGFTKLDYPKLCRIAPVILFAAAVSRA	150
30	orf130ng	LQNLPAAPLHLITLGGMTGGVMVWLTAGLWHSFGFTKLDYPKLCRIAVSILFASAVSRA	321
	orf130.pep	FLXNVNPFXXFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPE	193
	orf130ng	VLMNVNPIFFITVPEILTAAVFVLYLLTFVPIFRANAFTDDPE	364

An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

1	MNKFTHPMR	PFFVGA	AVLA	ILGALVFFHQ	PRRYHPAPPN	FLGTYAAGCI
51	RRFFDYRFVG	PDGFFRQPET	CRYFDGGVVA	CCGCFIAVFT	ATCRIFRRL	
101	LAGVAAVLRL	ADLARQHRT	LRSDVDTAAF	TVFQTAYAVS	GDNLLRAQV	
151	HLNMAAVMFV	SVRVSLLGT	ETLKECRLKD	PVFIPNVIYK	NIAITLLLHA	
201	AAELWLPAQT	AGFTALAVGF	ILLAKLRELH	HHELLRKHYV	RTYYLLQLFA	
251	AAGYLWTGAA	KLQNLPAAP	LHLITLGGMT	GGVMVWLT	GLWHSFGFTKL	
301	DYPKLCRIAV	SILFASAVSR	AVLMNVNPIF	FITVPEILTA	AVFMYLLTF	
351	VPIFRANAFT	DDPE*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

45	1	ATGCGCCCGT	TTTTCGTCGG	TGCGGCAGTA	CTTGCCATAC	TCGGTGCGTT
	51	GGTGTTTTTT	ATCAACCCCG	GCGCTATCAT	CCTGCACCGC	CAAATTTTCT
	101	TGGAACCTAT	GCTGCCGGCT	GCATACGGCG	GTTTTTTGAC	TACCGCTTTG
	151	TTGGACCGGA	CGGTTTTTTC	AGGCAACCTG	AAACCTGCCG	CTACTTTGAT
50	201	GGCGGTGTTG	TTGCTTGTG	CGGCTGTTTT	ATTGCCGTTT	TTACCGCAAC
	251	TTGCCGCATT	TTTCGTCGCC	GCCTATTGGC	TGGTGTTGCT	GCTGTTCTGC
	301	GCCTGGCTGA	TTTGGCTCGA	CCGCAACACC	GACAACTTCG	CTCTGTTGAT
	351	GTTACTTGCC	GCATTTACCG	TTTTTCAGAC	GGCCTATGCC	GTCAGCGCGC
	401	ATTTGAACTT	ACTGCGCGCG	CAAGTGCATT	TGAATATGGC	GGCGGTCATG
55	451	TTCGTATCCG	TCCGCGTCAG	CGTCCTTTTG	GGCACGGAAA	CCCTGAAAGA
	501	ATGCCGTCTG	AAAGACCCCG	TATTCATCCC	CAACGTATAT	TATAAAAACA
	551	TCGCCATCAC	CCTGCTGCTG	CACGCGCGCG	CCGAACCTTG	GCTGCGCGCG
	601	CAAACCGCGC	GTTTTACTGC	GCTTGCCGTC	GGCTTCATCC	TGCTCGCCAA
	651	GCTGCGCGAA	CTGCACCATC	ACGAACCTT	ACGCAACAC	TACGTCCGCA
	701	CTTATTACCT	GCTCCAGCTC	TTTGCCGCGC	CAGGTTATCT	GTGCACAGGC
60	751	GCGGCGAAAC	TGCAAAACCT	GCCCGCCTCC	GCGCCCTGCG	ACCTGATTAC
	801	CCTCGGCGGC	ATGACGGGTG	GCGTGATGAT	GGTGTGGCTG	ACTGCCGGAC
	851	TGTGGCACAG	CGGCTTTACC	AAACTCGACT	ACCCGAAACT	CTGCCGCATC

-465-

901 GCCGTCTCCA TCCTTTTCGC CTCCGCCGTT TCGCGCGCTG TTTTAATGAA
 951 CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG
 1001 TGTTTCATGCT TTACCTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACCGC
 1051 TTTACAGACG ATCCGGAATA A

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

1 MRPFFVGA AV LAILGALVFF INPGAAILHR QIFLEMLPA AYGGFLTTAL
 51 LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVLVLLFC
 101 AWLIWLD RNT DNFA LLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
 151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAAEWLWPA
 201 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYLLQL FAAAGYLWTG
 251 AAKLQNL PAS APLHLITLGG MTGGVMVWL TAGLWHS GFT KLDYPKLCRI
 301 AVSILFASAV SRAVL MNVNP IFFITVPEIL TAAVFMLYLL TFPVIFRANA
 351 FTDDPE*

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

15 orf130-1.pep MRPFFVGA AV LAILGALVFF INPGAIVLHRQIFLEMLPAA YGGFLTAALLDWTGFSGNL
 orf130ng-1 MRPFFVGA AV LAILGALVFF INPGA IILHRQIFLEMLPAA YGGFLTTALLDRTGFSGNL
 20 orf130-1.pep KPVATLMAALLAASAILPFSPQTASFFVAAYWLVLVLLFCARLIWLD RNT DNFA LLMLLA
 orf130ng-1 KPAATLMAV LLLVA AVLLPFLPQLAAFFVAAYWLVLVLLFCAWLIWLD RNT DNFA LLMLLA
 25 orf130-1.pep AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRS ILLGAEALKECRLKDPVFIPNIV
 orf130ng-1 AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRS VLLGTETLKECRLKDPVFIPNVI
 30 orf130-1.pep YKNIAITFLLLHAAAEWLWPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYLLQ
 orf130ng-1 YKNIAIT-LLLHAAAEWLWPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYLLQ
 35 orf130-1.pep LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMMGGVMVWL TAGLWHS GFT KLDYPKLCR
 orf130ng-1 LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMVWL TAGLWHS GFT KLDYPKLCR
 40 orf130-1.pep IAVPILFAAAVSRAFLMNVNPIFFITVPA ILTA AEFVLYLFTFIPIFRANAFTDDPEX
 orf130ng-1 IAVSILFASAVSRAVL MNVNP IFFITVPEILTA AEFMLYLLTFVPVIFRANAFTDDPEX

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

45 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
 101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
 151 GGCGGCGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
 201 CGGCAATAGT TCCGT CAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
 251 ACTTTTACAG GAAAATAGGG AAGTTGAAG C.TGCGGGCT GGATTGGCGT
 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGTA
 351 CTGCTTGGA AAG..

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
 51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSYFYRKIG KFEXCGLDWR
 101 TRDGKPLIET FKQGGFDCLE K..

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

55 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

-466-

5
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGA
351 CTGCTTGAA AAGCAGGGGT TGCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

10
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

20
orfl31a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
orfl31a MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
25
orfl31a.pep YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE
orfl31a YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE
30
orfl31a.pep K
orfl31a KQGLRRNGLSERVRWX
130

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

35
1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
51 TACGCTTGCA GGCTGCCGGT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
151 GCGGCGGAGA GTCCGCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
40
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA
351 TTGTTTGAAA AAGCAGGGGT TGCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

45
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQEGFDCLE KQGLRRNGLS ERVRW*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

50
orfl31a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLED
orfl31-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
orfl31a.pep YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE
orfl31-1 YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE
55
orfl31a.pep KQGLRRNGLSERVRWX
|||||

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

	orf131.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD	60
	orf131.ng	MEIRVIKYTATAALFAFTVAGCRLAGWYECLSLSGWCKPRKPAAIDFWDIGGESPLSLED	60
10	orf131.pep	YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFQGGFDCLE	120
	orf131.ng	YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE	120
15	orf131.pep	K 121	
	orf131.ng	KQGLRRNGLSERVRW 134	

20 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
 51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
 101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

	1	ATGGAAATTC	GGGTAATAAA	ATATACGGCA	ACGGCTGCGT	TGTTTGCAAT
25	51	TACGGTTGCA	GGCTGCCGGC	TGGCGGGGTG	GTATGAGTGT	TCGTCTTTGT
	101	CCGGCTGGTG	TAAGCCGAGA	AAACCTGCCG	CCATCGATTT	TTGGGATATT
	151	GGCGGCGAGA	GtcgcgtGTC	TTTAGAGGAC	TACGAGATAC	CGCTTTTCAGA
	201	CGGCAATCGT	TCCGTCAGGG	CAAACGAATA	TGAATCCGCG	CAAAAATCTT
	251	ACTTTTATAG	GAAAATAGGG	AAGTTTGAAG	CCTGCGGGTT	GGATTGGCGT
30	301	ACGCGTGACG	GCAAACCTTT	GGTTGAGAGG	TTCAAACAGG	AAGGTTTCGA
	351	CTGTTTGGAA	AAGCAGGGGT	TGCGGCGCAA	CGGCCTGTCC	GAGCGCGTCC
	401	GATGGTAA				

```

1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KOGLRRNGLS ERVRW*

```

```

40 orf131ng-1.pep MEIRVIKYTTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAADFDWDIGGESPLSLED
    orf131-1      MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAADFDWDIGGESPPSLGD

    orf131ng-1.pep YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
    orf131-1      YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

45 orf131ng-1.pep KQGLRRNGLSERVRWX
    orf131-1      KQGLRRNGLSERVRWX

```

50 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
    51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
  101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
  151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
  201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
  251  TGAACCTCGG CCTGCCTTAT ATTtCGGCC CGCAATGGCT GTCGGAACAA
  301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
  351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
  401  CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtCGGCG TTTCCGCCCG
  451  CCTGCCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTTcG
  501  TCATCGAAGC CGACGAATAC GACACCGCCT TTTtCGACAA ACgTtCTAAA
  551  TtCGTGCATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAAATCGA
  601  CCACGCCGAC ACTTGGGCGC GATACAGACc CAGTtCCACT
  651  ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
  701  CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGa
  751  AAAATTCGGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

  20      1  MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
      51  YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
     101  VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRFRFP
     151  PAANAAPRPE QPIAVFRHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
     201  PRRHLCRLGR DTDVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
     251  KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
    51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
   101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
   151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
   201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
   251  TGAACCTCGG CCTGCCTTAT ATTTCCGGCC CGCAATGGCT GTCGGAACAA
   301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
   351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
   401  CGGGCTTCCT TATTGGCGGC GTACCGGAAA ATTTCCGCGT TTTCCGCCCG
   451  CTGCCGCAA ACGCCGCGCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
   501  CATCGAAGCC GACGAATACG ACACCGCCTT TTTTCGACAA CGTTCATAAT
   551  TCGTGCATTA CCGTCCGCGT ACCGCGGTGT TGAACAATCT GGAATTCGAC
   601  CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCACATA
   651  CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
   701  AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGTGGAA
   751  AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
   801  CTCGTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
   851  ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTCGCCCC
   901  GCGCGTCATG TCGGTGTCGA TATTCAGACC GCCTGCGAAG CCTTGGGCGC
   951  GTTTAAAAAC GTCAAACGCC GGATGGAAAT CAAAGGCACG GCAAACGGCA
  1001  TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
  1051  ATTCAAGGTT TGCGCCAACG CGTCGGCGGC GCGCGCATCC TCGCGTCTCT
  1101  CGAACC GCGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
  1151  CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGCGGTG
  1201  GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
  1251  CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
  1301  TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
  1351  GGAAAGCTGC TGAAGCTTT GAGATAG

```

55 This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

      1  MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
     51  YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
    101  VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
    151  LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
    201  HADIFADLGA IQTFHYLVR TVPSEGLIVC NGRQOSLQDT LDKGCWTPVE
    251  KFGTEHGWQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
    301  ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPATAIETT

```

-469-

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLEKA DQVFCYAGGV
 401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH
 451 GKLEALR*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E. coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

Orf132: 4 IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLDEFK 63
 IHI+GI.GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +
 10 o457: 3 IHIIGIGGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61
 Orf132: 64 ADVYVIGNVAKRGMVDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTASML 123
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121
 15 Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143
 W+LE G PGF+IGGV G
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N. meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
25	orf132.pep	MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD					
	orf132a	MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD					
		10	20	30	40	50	60
30	orf132.pep	EFKADVYVIGNVAKRGMVDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA					
	orf132a	EFKADVYVIGNVAKRGMVDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA					
		70	80	90	100	110	120
35	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQPI-----AVFR					
	orf132a	SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQPI-----AVFR					
		130	140	150	160		
40	orf132.pep	HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRL					
	orf132a	HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRL					
		170	180	190	200	210	220
45	orf132a	KRSKFVHYRPRTAFLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGRQQLSD					
		180	190	200	210	220	230

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT
 51 TGCCGCCATT GCCAAAGAAG CAGGGTTTGA ANTCAGCGGT TGCGATGCGA
 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG
 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA
 201 CGTTATCGGC AATGTGCGCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC
 301 NTGCTGCACC ATCATTGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC
 351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGACTCGCAC
 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC
 451 CTGCCGCAAA CGCGCGGCGC AGACCGAAC AGCCAATCGC CGTTTTCGT
 501 CATTGAAGCC GACGAATACG ACACCGCGT TTTGACAAA CGCTCCAAAT
 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCACCA
 651 CCTCGTCCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGACGGC
 701 AGCAAAGCCT GCAAGACACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
 751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGATGG

-470-

5
10
1001
1051
1101
1151
1201
1251
1301
1351

```

801 CTCGTTTCGAC GTGTTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCTTGA
851 GTTTGATGGG CGGACACAAC CGCATGAACG CGCTCGCNGT CATCGCCGCC
901 GCGCGTCATG CCGGAGTNGA CATTCAGACG GCCTGCGAAG CCTTGAGCAC
951 GTTTAAAAAC GTCAAACGCC GCATGGAAAT CAAAGGCACG GCAAACGGTA
1001 TCACCGTTTA CGACGACTTC GCCCACCATC CGACCGCTAT CGAAACCACG
1051 ATTCAGGTT TCGCCAGCG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
1101 CGAACC GCGT TCCAATACGA TGAAGCTGGG TACGATGAAA GCCGCCCTGC
1151 CCGCAAGCCT CAAAGAAGCC GACCAAGTGT TCTGNTACGC CGGCGGCGCG
1201 GACTGGGACG TTGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGCACGT
1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTAAA AACGCCAAG
1301 CAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
1351 ACCAACTGC TGGACGCTTT GAGATAG

```

This encodes a protein having amino acid sequence <SEQ ID 870>:

15
20
401
451

```

1 MKHIHIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV
51 YEGFDTAQLD EFKADVIVIG NVAKRGM DVV EAILNRGLPY ISGPQWLAEN
101 XLHHHWX LGV AXTHGKT TTA SMLAWVLEYA GLAPGFXIGG VPENFSVSAR
151 LPQTPRODPN SQSPFFVIEA DEYDTAF FDK RSKFVHYRPR TAVLNNLEFD
201 HADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQOSLQDT LDKGCWTPVE
251 KFGTEHGWQA GEANADGSFD VLLDGKKAGH VAWSLMGGHN RMNALAVIAA
301 ARHAGVDIQT ACEALSTFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT
351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK AALPASLKEA DQVFX YAGGA
401 DWDVAEALAP LGGRLHVGKD FDAFVAEIVK NAEAGDHILV MSNGGFGGIH
451 TKLLDALR*

```

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25
30
35
40
45
50
55

```

orf132a.pep MKHIHIIGIGGTFMGGIAAIKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD
orf132-1 MKHIHIIGIGGTFMGGIAAIKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD

orf132a.pep EFKADVIVIGNVAKRGM DVVEAILNRGLPYISGPQWLAENXLHHHWX LGVAXTHGKT TTA
orf132-1 EFKADVIVIGNVAKRGM DVVEAILNLGLPYISGPQWLSENVLHHHWV LGVAGTHGKT TTA

orf132a.pep SMLAWVLEYAGLAPGFXIGGVPENFSVSARLPQTPRODPNSQSPPFFVIEADEYDTAF FDK
orf132-1 SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRODPNSQSPPFFVIEADEYDTAF FDK

orf132a.pep RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQOSLQDT
orf132-1 RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQOSLQDT

orf132a.pep LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKKAGH VAWSLMGGHN RMNALAVIAA
orf132-1 LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVK WDMGRHNR MNALAVIAA

orf132a.pep ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPTAIETT IQGLRQRVGG
orf132-1 ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPTAIETT IQGLRQRVGG

orf132a.pep ARILAVLEPR SNTMKLGTMKAALPASLKEADQVFX YAGGADWDVAEALAP LGGRLHVGKD
orf132-1 ARILAVLEPR SNTMKLGTMKSALPVSLKEADQVFX YAGGADWDVAEALAP LGGRLNVGKD

orf132a.pep FDAFVAEIVK NAEAGDHILV MSNGGFGGIHTKLLDALRX
orf132-1 FDAFVAEIVK NAEAGDHILV MSNGGFGGIH GKLLLEALRX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60

```

orf132.pep MKHIHIIGIGGTFMGGIAAIKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD 60
orf132ng MKHIHIIGIGGTFMGGIAAIKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAAQLE 60

```

-471-

	orf132.pep	EFKADVVYIGNVAKRGMVDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDVVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHRLRLGRDTPVPPRAHRTIRPHRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
15	orf132ng	FGQRLLDAGGKIRHRTRLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
20	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPKGFRFRFP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRHLRLGR	DTDPVPPRA	HRTIRPHRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRTRLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGGTTCAA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TCGATGCCGC	GCAATTGGAA	GAATTTCAAG	CCGATATTTA
	201	CGTCATCGGC	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
30	251	TGAACCGTGG	GCTGCCTTAT	ATTTCGGGCC	CGCAATGGCT	GGCTGAAAac
	301	GTGCTgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggaCGC	ACGGcaaaAac
	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCTT	GGAATATGCC	GGACTCGCGC
	401	CGGGCTTCCT	CATCGGCGGt	gtaccggaAA	ATTTCGGCGT	TTCGCGCCGC
	451	CTACCGCAAA	CGCCGCGTCA	AGACCCGAAC	AGCAAATCGC	CGTTTTTCGT
35	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCGAAT
	551	TCGTGCATTA	TCGCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	CTTGGGCGCG	ATACAGACCC	AGTTCACCA
	651	CCTCGTGCGC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTGGAA
40	751	AAATTGCGCA	CCGGACACGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTCGAC	GTATTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTCGCATGGG
	851	ATTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGTGGCC
	901	GCACGCCATG	CCGGAGTCGA	TGTTGAGACG	GCCTGCGAAG	CCTTGGGTGC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAAT	CAAAGGCACG	GCAAACGGCA
45	1001	TCACCGTTTA	CGACGATTTC	GCCCACCACC	CGACCGCCAT	CGAAACCACG
	1051	ATTCAGGTT	TGCGCCAACG	TGTCGGCGGC	GCGGCATCC	TCGCCGTCTT
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGCTACGC	CGGCGGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
50	1251	CGGTAAAGAT	TTCGATACCT	TCGTTGCCGA	AATTGTGAAA	AACGCCCGAA
	1301	CCGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACCTGC	TGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
55	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
	151	LPQTPRQDPN	SKSPFFVIEA	DEYDTAFFDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTFVE
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGHNN	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
	401	DWDVAEALAP	LGCRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSNGGFGGIIH
	451	TKLLDALR*				

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLE
   orf132-1      MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD

10 orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA
   orf132-1      EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA

15 orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGV PENFGVSARLPQT PRQDPNSKSPFFVIEADEYDTAFFDK
   orf132-1      SMLAWVLEYAGLAPGFLIGGV PENFGVSARLPQT PRQDPNSQSPFFVIEADEYDTAFFDK

20 orf132ng-1.pep RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHHLV R T V P S E G L I V C N G Q Q Q S L Q D T
   orf132-1      RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHYLV R T V P S E G L I V C N G R Q Q S L Q D T

25 orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSGFDVLLDGKKAGHVAWDL MGGHNRMNALAVIAA
   orf132-1      LDKGCWTPVEKFGTEHGWQAGEANADGSGFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA

30 orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
   orf132-1      ARHVGVDIQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

35 orf132ng-1.pep ARILAVLEPRSNMTKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRRLRVGKD
   orf132-1      ARILAVLEPRSNMTKLGTMKSALPVSLKEADQVFCYAGGVWDVAEALAPLGGRNLNVGKD

40 orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
   orf132-1      FDAFVAEIVKNAEVDGHILVMSNGGFGGIHGKLLLEALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

```

35  pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
   ORF o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
   in fbp-pmba intergenic region [Escherichia coli] Length = 457
   Score = 474 bits (1207), Expect = e-133
   Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

40  Query: 22 KEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
   ++ G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE
   Sbjct: 21 RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

45  Query: 82 AILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141
   A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV
   Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCQGYKPGFVIGGV 139

50  Query: 142 PENFGVSARLPQT PRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPTAVLNNLEFDH 201
   P NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH
   Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

55  Query: 202 ADIFADLGAIQTQFHHLV R T V P S E G L I V C N G Q Q Q S L Q D T L D K G C W T P V E K F G T G H G W Q I G 261
   ADIF DL AIQ QFHHLV R VP +G I+ +L+ T+ GCW+ E G WQ
   Sbjct: 191 ADIFDDLKAIQKQFHHLV R I V P G Q G R I I W P E N D I N L K Q T M A M G C W S E Q E L V G E Q G H W Q A K 250

60  Query: 262 EVNADGS-FDVLLDGKKAGHVAWDL MGGHNRMNALAVIAAARHAGVDVQTACEALGAFKN 320
   ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N
   Sbjct: 251 KLTTDASEWEVLLDGEKVGEVKWSLVGEHNMHNGLMAIAAARHVGVA P A D A A N A L G S F I N 310

65  Query: 321 VKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMTKLGTM 379
   +RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNMTMK+G
   Sbjct: 311 ARRLLELRGEANGVTVYDDFAHHPTAILATLAALRGKVGGTARIIVLEPRSNMTKMGIC 370

70  Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRRLRVGKDFDTFVAEIVKNARTGDHI 438
   K L SL AD+VF W VAE D DT +VK A+ GDHI
   Sbjct: 371 KDDLAPSLGRADEVFLLPAPHPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430

75  Query: 439 LVMSNGGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L
Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These
10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1   . . CCGGGCTATT  ACGGCTCGGA  TGACGAATTT  AAGCGGGCAT  TCGGAGAAAA
51  CTGCGCCGACA  TmCAAGAAAC  ATTGCAACCG  GAGCTGCGGG  ATTTATGAAC
15  101 CCGTATTGAA  AAAATACGGC  AAAAAGCGCG  CCAACAACCA  TTCGGTCAGC
151 ATTAGTGCGG  ACTTCGGCGA  TTATTTCATG  CCGTTCGCCA  GCTATTCGCG
201 CACACACCGT  ATGCCCAACA  TCCAAGAAAT  GTATTTTCC  CAAATCGGCG
251 ACTCCGGCGT  TCACACCGCC  TTAAGAACAG  AGCGCGCAAA  CACTTGGCAA
301 TTTGGCTTCr  ATACCTATAA  AAAAGGATTG  TTAACAACAG  ATGATACATT
20  351 AGGATTAAAA  CTGGTCGGCT  ACCGCAGCCG  CATCGACAAC  TACATCCACA
401 ACGTTTACGG  GAAATGGTGG  GATTTGAACG  GGGATATTC  GAGCTGGGTC
451 AGCAGCACCG  GGCTTGCCTA  CACCATCCAA  CATCGCrATT  TCAGAGACAA
501 AGTGCATCAA  nnnnnnnnnn  nnnnnnnnnn  nnnnnTACGAT  TATGGGCGTT
551 TTTTCACCAA  CCTTCTTAC  GCCTATCAAA  AAAGCACGCA  ACCGACCAAC
25  601 TTCAGCGATG  CGAGCGAATC  GCCCAACAAT  GCGTCCAAAG  AAGACCAACT
651 CAAACAAGGT  TATGGGTTGA  GCAGGGTTTC  CGCCCTGCCG  CGAGATTACG
701 GACGTTTGGG  AGTCGGTACG  CGCTGGTTGG  GCAACAAACT  GACTTTGGGC
751 GGCGCGATGC  GCTATTTCCG  CAAGAGCATC  CGCGCGACGG  CTGAAGAACG
801 CTATATCGAC  GGCACCAACG  GGGGAAATAC  CAGCAATTC  CGGCAACTGG
30  851 GCAAGCGTTC  CATCAAACAA  ACCGAACTC  TTGCCCGCCA  GCCTTTGATT
901 TTWgATTTTa  ACGCCGCTTA  CGAGCCGAAG  AAAAACCTTA  TTTTCCGCGC
951 CGAAGTCAAA  AATCTGTTCG  ACAGGCGTTA  TATCGATCCG  CTCGATGCGG
1001 GCAATGATGC  GGCAAC .GAG  CGTTATTACA  GCTCGTTCCA  CCCGAAAGAC
1051 AAGGACrrAG  ACGTAACGTG  TAATGCTGAT  AAAACGTTGT  GCaACGGCAA
35  1101 ATACGGCGGC  ACAAGCAAAA  GCGTATTGAC  CAATTTTGCA  CGCGGACGCA
1151 CCTTTTTgAT  GACGATGAGC  TACAAGTTTT  AA

```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1   . . PGYYGSDDEF  KRAFGENSPT  XXXKHCNRSCG  IYEPVLKKYG  KKRANNHSVS
51  ISADFGDYFM  PFASYSRTHR  MPNIQEMYFS  QIGDSGVHTA  LKPERANTWQ
40  101 FGFXTYKKGL  LKQDDTLGLK  LVGYRSRIDN  YIHNVYGKWW  DLNGDIPSWV
151 SSTGLAYTIQ  HRXFXDKVHQ  XXXXXXXXYD  YGREFTNLSY  AYQKSTQPTN
201 FSDASESPNN  ASKEDQLKQG  YGLSRVSALP  RDYGRLEVGT  RWLGNKLTLG
251 GAMRYFGKSI  RATAEERYID  GTNGGNTSNF  RQLGKRSIKQ  TETLARQPLI
301 XDFNAAYEPK  KNLI FRAEVK  NLFDRRYIDP  LDAGNDAAXE  RYSSFFDPKD
45  351 KDXDVT CNAD  KTL CN GKYGG  TSKSVLTNFA  RGRTFLMTMS  YKF*

```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1   GAGGCGCAGA  TACAGGTTTT  GGAAGATGTG  CACGTCAAGG  CGAAGCGCGT
51  ACCGAAAGAC  AAAAAAGTGT  TTACCGATGC  GCGTGCCGTA  TCGACCCGTC
50  101 AGGATATATT  CAAATCCAGC  GAAAACCTCG  ACAACATCGT  ACGCAGCATC
151 CCCGGTGCGT  TTACACAGCA  AGATAAAGC  TCGGGCATTG  TGTCTTTGAA
201 TATTCGCGGC  GACAGCGGGT  TCGGGCGGGT  CAATACGATG  GTGGACGGCA
251 TCACGCAGAC  CTTTTATTTC  ACTTCTACCG  ATGCGGGCAG  GGCAGGCGGT

```

-474-

301	TCATCTCAAT	TCGGTGCATC	TGTCGACAGC	AATTTTATTG	CCGGACTGGA
351	TGTCGTCAAA	GGCAGCTTCA	GCGGCTCGGC	AGGCATCAAC	AGCCTTGCCG
401	GTTTCGGCGAA	TCTGCGGACT	TTAGGCGTGG	ATGACGTCGT	TCAGGGCAAT
451	AATACCTACG	GCCTGCTGCT	AAAAGGTCTG	ACCGGCACCA	ATTCAACCAA
5	501	AGGTAATGCG	ATGGCGGCGA	TAGGTGCGCG	CAAATGGCTG
	551	CATCTGTCGG	TGTGCTTTAC	GGGCACAGCA	GGCGCAGCGT
	601	TACCGCGTGG	GCGGCGGCGG	GCAGCACATC	GGAAATTTTG
	651	TTTGAACGG	CGCAAGCAGC	GATATTTTGT	ACAAGAGGGT
	701	TCAATTCCGA	CAGCGGAAAA	TGGGAGCGGG	ATTTACAAAG
10	751	AAATACAAGC	CGTATAAAAA	TTACAACAAC	CAAGAACTAC
	801	CGAAGAGCAT	GACAAAAGCT	GGCGGGAAAA	CCTg . CaCCG
	851	TTACCCCCAT	CGATCCGTCC	AGCCTGAAGC	AGCAGTCGGC
	901	TTTAAATTGG	AATACGACGG	CGTATTCAAT	AAATACACGG
	951	CGATTTAAAC	ACCAAAATCG	GCAGCCGCAA	AATCATCAAC
15	1001	AGTTCAATTA	CGGTTTGTCT	TTGAACCCGT	ATACCAACCT
	1051	GCAGCCTACA	ATTCGGGCAG	GCAGAAATAT	CCGAAAGGGT
	1101	AGGCTGGGGG	CTTTTAAAGG	ATTTTGAAAC	CTACAACAAC
	1151	TCGACCTCAA	CAACACCGCC	ACCTTCCGGC	TGCCCCGCGA
	1201	CAAACCACTT	TGGGCTTCAA	TTATTTCCAC	AACGAATACG
20	1251	CTTTCCTGAA	GAATTGGGGC	TGTTTTTCGA	CGGTCTGAT
	1301	GGCTTTATTC	CTATTTGGGG	CGGTTTAAAG	GCGATAAAGG
	1351	CAAAAATCAA	CCATTGTCCA	ACCGGCCGGC	AGCCAATATT
	1401	CTACTTCGAT	GCCGCGCTCA	AAAAAGACAT	TTACCGCTTA
	1451	CCAATACCGT	CGGTACCGT	TTCGGCGGCG	AATATACGGG
25	1501	TCGGATGACG	AATTTAAGCG	GGCATTGCGA	GAAAACTCGC
	1551	GAAACATTGC	AACCGGAGCT	GCGGGATTTA	TGAACCCGTA
	1601	ACGGCAAAAA	GCGCGCCAAC	AACCATTCCG	TCAGCATTAG
	1651	GGCGATTATT	TCATGCCGTT	CGCCAGCTAT	TCGCGCACAC
	1701	CAACATCCAA	GAAATGTATT	TTTCCCAAAT	CGGCGACTCC
30	1751	CGGCCTTAAA	ACCAGAGCGC	GCAAACACTT	GGCAATTTGG
	1801	TATAAAAAAG	GATTGTAAAA	ACAAGATGAT	ACATTAGGAT
	1851	CGGTACCGC	AGCCGCATCG	ACAACACTAC	CCACAACGTT
	1901	GGTGGGATTT	GAACGGGGAT	ATTCCGAGCT	GGGTGAGCAG
	1951	GCCTACACCA	TCCAACATCG	CAATTTCAAA	GACAAAGTGC
35	2001	TTTTGAGTTG	GAGCTGAATT	ACGATTATGG	GCGTTTTTTC
	2051	CTTACGCCTA	TCAAAAAAGC	ACGCAACCGA	CCAACCTCAG
	2101	GAATCGCCCA	ACAATGCGTC	CAAAGAAGAC	CAACTCAAAC
	2151	GTTGAGCAGG	GTTTCCGCCC	TGCCGCGAGA	TTACGGACGT
	2201	GTACGCGCTG	GTTGGGCAAC	AACTGACTT	TGGGCGGCGC
40	2251	TTCGGCAAGA	GCATCCGCGC	GACGGCTGAA	GAACGCTATA
	2301	CAACGGGGGA	AATACCAGCA	ATTTCCGGCA	ACTGGGCAAG
	2351	AACAAACCGA	AACTCTTGCC	CGCCAGCCTT	TGATTTTTGA
	2401	GCTTACGAGC	CGAAGAAAAA	CCTTATTTTC	CGCGCCGAAG
	2451	GTTTCGACAGG	CGTTATATCG	ATCCGCTCGA	TGCGGGCAAT
45	2501	CGCAGCGTTA	TTACAGCTCG	TTCGACCCGA	AAGACAAGGA
	2551	ACGTGTAATG	CTGATAAAAC	GTTGTGCAAC	GGCAAATACG
	2601	CAAAAGCGTA	TTGACCAATT	TGACACGCGG	ACGCACCTTT
	2651	TGAGCTACAA	GTTTTAA		

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50	1	EAQIQVLEDV	HVKAKRVPKD	KKVFTDARAV	STRODIFKSS	ENLDNIVRSI
	51	PGAFTQODKS	SGIVSLNIRG	DSGFGRVNTM	VDGITQTFYS	TSTDAGRAGG
	101	SSQFGASVDS	NFIAGLDVVK	GSFSGSAGIN	SLAGSANLRT	LGVDVVQGN
	151	NTYGLLLKGL	TGTNSTKGNA	MAAIGARKWL	ESGASVGVLY	GHSRRSVAQN
55	201	YRVGGGGQHI	GNFGAEYLER	RKQRYFVQEG	ALKFNSDSGK	WERDLQRQW
	251	KYKPYKNNYN	QELQYIEEH	DKSWRENLXP	QYDITPIDPS	SLKQOSAGNL
	301	FKLEYDGVFN	KYTAQFRDLN	TKIGSRKIIN	RNYQFNYGLS	LNPYTNLNL
	351	AAYNSSGRQKY	PKGSKFTGWG	LLKDFETYN	AKILDNLNTA	TFRLPRETEL
	401	QTTLGFNYFH	NEYGKNRFFE	ELGLFFDGP	QDNGLYSYLG	RFKWDKGLLP
	451	QKSTIVQPAG	SOYFNTFFFD	AALKKDIYRL	NYSTNTVGYR	FGGEYTGYYG
60	501	SDDEFKRAFG	ENSPYKHKHC	NRSCGIYEPV	LKKYGKKRAN	NHSVSIADNF
	551	GDYFMPFASY	SRTHRMPNIQ	EMYFSQIGDS	GVHTALKPER	ANTWQFGFNT
	601	YKKGLLKQDD	TLGLKLGVYR	SRIDNYIHNV	YGKWDDLNGD	IPSWDSTGL
	651	AYTIQHRNFK	DKVHKHGFEL	ELNYDYGRFF	TNLSYAYQKS	TQPTNFSNAS
	701	ESPNNASKED	QLKQGYGLSR	VSLPRDYGR	LEVGTWRLGN	KLTLLGAMRY
65	751	FGKSIRATAE	ERYIDGTNGG	NTSNFRQLGK	RSIKQTETLA	RQPLIFDFYA
	801	AYEPKKNLIF	RAEVKNLFDR	RYIDPLDAGN	DAATQRYSS	FDPKDKDEDV
	851	TCNADKTLN	KGYYGTSKSV	LTNFARGRTF	LMTMSYKF*	

Computer analysis of this amino acid sequence gave the following results:

Homology with with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

```

5   Orf133: 31 IYEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
    HI121: 563 INEPILHKSGHKAFNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFFSQVSNAGVNTA 622

10  Orf133: 91 LKPERANTWQFGFXYTKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150
    HI121: 623 LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYIHNVYGWW--RDGMPTWA 680

15  Orf133: 151 SSTGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDASESPNN 210
    HI121: 681 ESNGFKYTIAHQNYKPIVKKSGVELEINYDMGRFFANVSYAYQRTNQPTNYADASPRPNN 740

20  Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYID 270
    HI121: 741 ASQEDILKQGYGLSRVSMPLPKDYGRLELGTWFDQKLTGLAARYYGKSKRATIEEYIN 800

25  Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDP 330
    HI121: 801 GSR-FKKNLRLRENYAVKKTEDIKKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDP 859

30  Orf133: 331 LDAGNDAAXERYSSFDPKDKDXDVTGNADKTLNKGKYGGSKSVLTNFARGRTFLMTMS 390
    HI121: 860 LDAGNDAASQRYSSSL-----NNSIECAQDSSAC----GGSDKTVLYNFARGRTYILSLN 910

    Orf133: 391 YKF 393
    HI121: 911 YKF 913

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N. meningitidis*:

```

35  orf133.pep          10      20      30
    PGYYGSDDEFKRAFGENSPTXKKHNCNRSCGI
    orf133a      FYFDAALKKDIYRLNYSTNTVGYRFGGXYTGYYXSDDEFKRAFGENSPTYXKHNCQSCGI
    450      460      470      480      490      500

40  orf133.pep          40      50      60      70      80      90
    YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
    orf133a      YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
    510      520      530      540      550      560

45  orf133.pep          100     110     120     130     140     150
    KPERANTWQFGFXYTKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS
    orf133a      KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVS
    570      580      590      600      610      620

50  orf133.pep          160     170     180     190     200     210
    STGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDASESPNNA
    orf133a      STGLAYTIQHRNFKDKVHKHGFELNLDYXRFFTNLSYAYQKSTQPTNFSDASESPNNA
    630      640      650      660      670      680

55  orf133.pep          220     230     240     250     260     270
    SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDG
    orf133a      SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDX
    690      700      710      720      730      740

60  orf133.pep          280     290     300     310     320     330
    TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL
    orf133a      TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL

```

A partial ORF133a nucleotide sequence <SEQ ID 879> is:

	1	AAAGACAAAA	AAGTGTTTAC	CGATGCGCGT	GCCGTATCGA	CCCGTCAGGA
	51	TATATTCAAA	TCCANCGAAA	ACCTCGACAA	CATCGTACGC	ANCATCCCCG
20	101	GTGCGTTTAC	ACANCAANAT	AAAAGCTCGG	GCNTTGTGTC	TTTGAATATT
	151	CGCNGCGACA	CGGGGTTCGG	CTGGGTCAAT	ACNATGGTNG	ACGGCATCAC
	201	NCANACCTTT	TATTCGACTT	CTACCGATGC	GGGCAGGGCA	GGCGGTTTCAT
	251	CTCAATTCCG	TGCATCTGTC	GACAGCAATT	TTATNGCCGG	ACTGGATGTC
	301	GTCAAAGGCA	GCTTCAGCGG	CTCGGCAGGC	ATCAACAGCC	TTGCGGTTTC
25	351	GGCGAATCTG	CGGACTTTAN	GGCTGGATGA	TTCGTTTCAG	GGCAATANTA
	401	CNTACGGCCT	GCTGCTAAAA	GGTCTGACCG	GCACCAATTC	AACCAAAGGT
	451	AATGCGATGG	CGGCGATAGG	TGCGCGCAAA	TGGCTGGAAA	GCGGAGCATC
	501	TGTCGGTGTG	CTTTACGGGC	ACAGCAGGCG	CAGCGTGGCG	CAAAATATCC
	551	GCGTGGGCGG	CTGGGGGCAG	CACATCGGAA	ATTTTGGCGC	GGAATATCTG
30	601	GAACGACGCA	AGCAACGATA	TTTTGAGCAA	GAAGGCGGGT	TGAAATTCAA
	651	TTCCAAACAG	GGAAAAATGG	AGCGGGATTT	CCAAAAGTCG	TACTGGAAAA
	701	CCAAGTGGTA	TCAAAAATAC	GATGCCCCCT	AAGAAGTACA	AAAATACATC
	751	GAAGGTCTAG	TAAAAAGCTG	GCGGGAAAAAC	CTGGCGCCGC	AATACGCAT
	801	CACCCCATC	GATCCGTCCA	GCCTGAAGCN	GCAGTCGGGA	GGCAACCTGT
35	851	TTAAATTGGA	ATACGACGGC	GTATTCAATA	AATACACGGC	GCAATTTTCG
	901	GATTTAAACA	CCAAAATCGG	CAGCCGCAAA	ATCATCAACC	GCAATTATCA
	951	ATTTCAATTAC	GGTTTGTCTT	TGAACCCGTA	TACCAACCTC	AATCTGACCG
	1001	CAGCCTACAA	TTCGGGCAGG	CAGAAATATC	CGAAAGGGTC	GAAGTTTACA
	1051	GGCTGGGGGG	TTTTNAAAGA	TTTTGAAACC	TACAACAACG	CAAAAATCTT
40	1101	CGACCTCANC	AACACCTCCA	CCTTCCGGCT	GCCCCGTGAA	ACCGAGTTGC
	1151	AAACCACTTT	GGGCTTCAAT	TATTTCCACA	ACGAATACGG	CAAAAACCGC
	1201	TTTCCTGAAG	AATTGGGGCT	GTTTTTCGAC	GGTCCGGATC	ANGACAACGG
	1251	GCTTTATTCC	TATTTGGGGC	GGTTTAAGGG	CGATAAAGGG	CTGCTGCCCC
	1301	AAAAATCAAC	CATTGTCCAA	CCGGCCGGCA	GGCAATATTT	CACACGTTTC
45	1351	TACTTCGATG	CCGCGCTCAA	AAAAGACATT	TACCGCTTAA	ACTACAGCAC
	1401	CAATACCGTC	GGCTACCGTT	TCGGCGGCNA	ATATACGGGC	TATTACNGCT
	1451	CGGATGACGA	ATTTAAGCGG	GCATTCCGAG	AAAATCGCC	GACATACANG
	1501	AAACATTGCA	ACCAGAGCTG	CGGAATTTAT	GAACCCGTAT	TGAAAAAATA
	1551	CGGCAAAAAG	CGCGCCAACA	ACCATTCCGT	CAGATTAGT	GCGGACTTCC
50	1601	GCGATTATTT	CATGCCGTTT	GCCAGCTATT	CGCGCACACA	CCGTATGCCC
	1651	AACATCCAAG	AAATGTATTT	TTCCCAATAT	GGCGACTCCG	CGGTTACAC
	1701	CGCCTTAAAA	CCAGAGCGCG	CAAAACCTTG	GCAATTTGGC	TTCAATACCT
	1751	ATAAAAAAGG	ATTGTTAAAA	CAAGATGATA	TATTAGGATT	AAAACTGGTC
55	1801	GGTACCCGCA	GCCGCATCGA	CNACTACATC	CACAACGTTT	ACGGGAAATG
	1851	GTGGGATTTG	AACGGGAATA	TTCCGAGCTG	GGTCAGCAGC	ACCGGGCTTG
	1901	CCTACACCAT	CCAACACCGC	AATTTCAAAG	ACAAAGTGCA	CAAACACGGT
	1951	TTTGAGTTGG	AGCTGAATTA	CGATTATNGG	CGTTTTTTCA	CCAACTTTTC
	2001	TTACGCCTAT	CAAAAAAGCA	CGCAACCGAC	CAACTTCAGC	GATGCGAGCG
	2051	AATCGCCCAA	CAATGCGTCC	AAAGAAGACC	AACTCAAAAC	AGGTTATGGG
60	2101	TTGAGCAGGG	TTTCCGCCCT	GCCGCGAGAT	TACGGACGTT	TGGAAGTCGG
	2151	TACGCGCTGG	TTGGGCAACA	AACTGACTTT	GGGCGCGCG	ATGCGCTATT
	2201	TCGGCAAGAG	CATCCGCGCG	ACGGCTGAAG	AACGCTATAT	TCAGCNCACC
	2251	AATGGGGNAN	NTACCAGCAA	TTTCCGGCAA	CTGGGCAAGC	GTTCCATCAN
	2301	ACAAACCGAA	ACCCTTGCCC	GCCAGCCTTT	GATTTTTTGT	TTNTACGCCG
	2351	CTTACGAGCC	GAAGAAAAAN	CTTATTTTCC	GCGCCGAAGT	CAAAAAATCT
65	2401	TTGCACAGGC	GTTATATCGA	TCCGCTCGAT	GCGGGCAATG	ATGCGGCAAC
	2451	GCAGCGTTAT	TACAGTTCTG	TCGACCCGAA	AGACAAGGAC	GAAGAAGTAA
	2501	CGTGTAATGA	TGATAACACG	TTATGCAACG	GCAAATACGG	CGGCACAAGC
	2551	AAAAGCGTAT	TGACCAATTT	TGCACGCGGA	CNCACCTTTT	TGATAACGAT
	2601	GAGCTACAAG	TTTTTAA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

      1 KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI
    51 RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
   101 VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
    5 151 NAMAAGARK WLESGASVGV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
   201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWKYQKY DAPQELQKYI
   251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR
   301 DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKET
   351 GWGLXKDFET YNNAKILDIX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR
  10 401 FPEELGLFFD GPDNDGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
   451 YFDAALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX
   501 KHCNQSCGIY EPVLKKYGKK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP
   551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLV
  15 601 GYRSRIDXYI HNVYKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
   651 FELELNIDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG
   701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
   751 NGXXTSNFRQ LKRSIXQTE TLARQPLIFD XYAAEYPPKKX LIFRAEVKNL
   801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNT LCNGKYGGTS
   851 KSVLTNFARG XTFLITMSYK F*

```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

                                     10      20      30      40
orfl33a.pep                      KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS
25 orfl33-1                      EAQIQVLEDVHVKAARVPKDKKVFTDARAVSTRQDIFKSSENLDNIVRSIPGAFTQQDKS
                                     10      20      30      40      50      60
orfl33a.pep                      SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGSSQFGASVDSNFXAGLDVVK
30 orfl33-1                      SGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGSSQFGASVDSNFIAGLDVVK
                                     70      80      90      100      110      120
orfl33a.pep                      GSFGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAGARKWL
35 orfl33-1                      GSFGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                     130      140      150      160      170      180
orfl33a.pep                      ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
40 orfl33-1                      ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK
                                     190      200      210      220      230      240
orfl33a.pep                      WERDFQKSYWKTWKYQKYDAPQELQKYIEGHDKSWRENLAPOYDITPIDPSSLKXQSAGN
45 orfl33-1                      WERDLQRQWQKYPYKNYNN-QELQKYIEEHDKSWRENLPQYDITPIDPSSLKQQSAGN
                                     250      260      270      280      290
orfl33a.pep                      LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
50 orfl33-1                      LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
                                     300      310      320      330      340      350
orfl33a.pep                      YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTTLGFNHYFHNEYGKNRFP
55 orfl33-1                      YPKGSKFTGWGLLKDFETYNNAKILDINNTATFRLPRETELQTTLGFNHYFHNEYGKNRFP
                                     360      370      380      390      400      410
orfl33a.pep                      EELGLFFDGPDPXNDGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
60 orfl33-1                      EELGLFFDGPDPQDNDGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                     420      430      440      450      460      470
orfl33a.pep                      EELGLFFDGPDPXNDGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
65 orfl33-1                      EELGLFFDGPDPQDNDGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR

```

-478-

		470	480	490	500	510	520
	orf133a.pep	LN	Y	S	T	N	T
	orf133-1	LN	Y	S	T	N	T
5		480	490	500	510	520	530
	orf133a.pep	NN	H	S	V	S	I
10	orf133-1	NN	H	S	V	S	I
		540	550	560	570	580	590
	orf133a.pep	TY	K	K	G	L	L
15	orf133-1	TY	K	K	G	L	L
		600	610	620	630	640	650
	orf133a.pep	K	D	K	V	H	K
20	orf133-1	K	D	K	V	H	K
		660	670	680	690	700	710
	orf133a.pep	R	V	S	A	L	P
25	orf133-1	R	V	S	A	L	P
		720	730	740	750	760	770
	orf133a.pep	K	R	S	I	X	Q
30	orf133-1	K	R	S	I	X	Q
		780	790	800	810	820	830
	orf133a.pep	S	F	D	P	K	D
35	orf133-1	S	F	D	P	K	D
		840	850	860	870	880	

Homology with a predicted ORF from *N.gonorrhoeae*ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*45 *gonorrhoeae*:

	orf133.pep	PGYYSDDDEFKRAFGENSPTXKKHCNRSCGI	31
	orf133ng	FYFDAALKKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
50	orf133.pep	YEPVLKKYGGKRANNHSVSIADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGGKRANNHSVSIADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQFGFXYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
60	orf133.pep	STGLAYTIQHRXFXDKVHQXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	211
	orf133ng	STGLAYTIRHRNFKDKVHKHGFLELNYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRLGNKLTGGAMRYFGKSIRATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRLGNKLTGGAMRYFGKSIRATAEERYIDG	800
65	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLI FRAEVKNLFDRRYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKNLI FRAEVKNLFDRRYIDPL	860

orf133.pep DAGNDAAXERYSSFDPKDKDXDVTCTNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY 391
 |||||:::||||| |||||
orf133ng DAGNDAATQRYSSFDPKDKDEDVTCNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY 920
5 orf133.pep KF 393
 ||
 orf133ng KF 922

10 The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a
protein having amino acid sequence <SEQ ID 882>:

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAIQIVL EDVHVKAERV
51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGSSSQFGAS VDSNFIAGLD
151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
15 201 GNAMAAIGAR KWLESASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
251 LERRKQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY
301 IEEHDKSWRE NLAPQYDITP IDPSGLKQOS AGNLLNLEYD GVFNKYTAQF
351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKF
401 TGWGLLKDFE TYNNAKILD NNTATFRLPR ETELQTTLGF NYFHNEYGKN
20 451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYYGSENEFK RAFGENSPAY
551 KEHCDPSCGL YEPVLKKYK KRANNHSVSI SADFGDYFMP FAGYSRTHRM
601 PNIQEMYFSQ IGD SGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
651 VGYRSRIDNY IHNVYGKWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH
25 701 GFELELNIDY GRFFTNL SYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
751 GLSRVSALPR DYGRLEVGT WLGKLTLLGG AMRYFGKSIR ATAERYIDG
801 TNGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAEYPPK NLI FRAEVKN
851 LFD RRYIDPL DAGNDAATQ YSSFDPKDK DEDVTCNADK TLCNGKYGGT
901 SKSVLTNFAR GRTFLMTMSY KF*

30 A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

1 ATGAGATCTT CTTTCCGGTT GAAGCCGATT TGT TTTTATC TTATGGGTGT
51 TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGGCAGCG
101 AGGCGCAGAT ACAGGTTTGT GAAGATGTGC ACGTCAAGGC GAAGCGCGTA
151 CCGAAAGACA AAAAAGTGTT TACCGATGCG CGTGCCGTAT CGACCCGTca
35 201 gGATGTGTTC AAATCCGGCG AAAACCTCGA CAACATCGTA CGCAGCATAC
251 CCGGTGCGTT TACACAGCAA GATAAAAGCT CGGGCATGTG GTCTTTGAAT
301 ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT
351 CACGCAGACC TTTTATTCTGA CTTCTACCGA TGCGGGCAGG GCAGGCGGTT
401 CATCTCAATT CCGGTGCATCT GTCGACAGCA ATTTTATTGC CGGACTGGAT
40 451 GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG
501 TTCGGCGAAT CTGCGGACTT TAGGCGTGGA TGACGTCGTT CAGGGCAATA
551 ATACCTACGG CCTGCTGCTA AAAGGTCTGA CCGGCACCAA TTCAACCAAA
601 GGTAATGCGA TGGCGGCGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC
651 GTCTGTCCGT GTGCTTTACG GGCACAGCAG GCGCGGCGGT GCGCAAAATT
45 701 ACCGCGTGGG CGGCGGCGGG CAGCACATCG GAAATTTTGG TGAAGAATAT
751 CTGGAACGGC GCAAACAGCA ATATTTTGTA CAAGAGGGTG GTTTGAAATT
801 CAATGCCGGC AGCGGAAAAT GGAACGGGA TTTGCAAAGG CAATCTGGA
851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCAAGAAGT GCAAAAATAC
901 ATCGAAGAGC ATGATAAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA
50 951 CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG GCAGGCAATC
1001 TGT TTA AATT GGAATACGAC GGCGTATTCA ATAAATACAC GCGCAATTT
1051 CGCGATTTAA ACACCAGAAT CGGCAGCCGC AAAATCATCA ACCGCAATTA
1101 TCAATTCAAT TACGGTTTGT CTTTGAACCC GTATACCAAC CTCAATCTGA
1151 CCGCAGCCTA CAATTCCGGG AGGCAGAAAT ATCCGAAAGG GGCGAAGTTT
55 1201 ACAGGCTGGG GGCTTTTAA AGATTTTGAA ACCTACAACA ACGCGAAAAT
1251 CCTCGACCTC AACAAACCCG CCACCTCCG GCTGCCCCG GAAACCGAGT
1301 TGCAAACCAC TTTGGGCTTC AATTATTTCC ACAACGAATA CGGCAAAAAC
1351 CGCTTTCCTG AAGAATTGGG GCTGTTTTTC GACGGTCTTG ATCAGGACAA
1401 CCGGCTTTAT TCCTATTTGG GGCGGTTTAA GGGCGATAAA GGGCTGTTGC
60 1451 CTCAAAAATC AACCAATTGTC CAACCGGCCG GCAGCCAATA TTTCAACACG
1501 TTCTACTTCG ATGCCGCGCT CAAAAAGAC ATTTACCGCT TAAACTACAG
1551 CACCAATGCA ATCAACTACC GTTTCGGCGG CGAATATACG GGCTATTACG
1601 GCTCGGAAAA CGAATTTAAG CGGGCATTTCG GAGAAAACTC GCCGGCATAC
1651 AAGGAACATT GCGACCCGAG CTGCGGGCTT TATGAACCCG TATTGAAAAA
65 1701 ATACGGCAAA AAGCGGCCCA ACAACCATTC GGTGAGCATT AGTGCGGACT
1751 TCGGCGATTA TTTATGCGCG TTCGCGCGCT ATTCGCGCAC ACACCGTATG

-480-

5
10
15
20

1801	CCCAACATCC	AAGAAATGTA	TTTTTCCCAA	ATCGGCGACT	CCGGCGTTCA
1851	CACCGCCTTA	AAACCAGAGC	GCGCAAACAC	TTGGCAATTT	GGCTTCAATA
1901	CCTATAAAAA	AGGATTGTTA	AAACAAGATG	ATATATTAGG	ATTGAAACTG
1951	GTCGGCTACC	GCAGCCGCAT	TGACAACTAC	ATCCACAACG	TTTACGGGAA
2001	ATGGTGGGAT	TTGAACGGGG	ATATTCCGAG	CTGGGTCGGC	AGCACC GGCG
2051	TTGCCTACAC	CATCCGACAC	CGCAATTTCA	AAGACAAAGT	GCACAAACAC
2101	GGTTTTGAGC	TGGAGCTGAA	TTACGATTAT	GGGCGTTTTT	TCACCAACCT
2151	TTCTTACGCC	TATCAAAAAA	GCACGCAACC	GACCAATTTT	AGCGATGCGA
2201	GCGAATCGCC	CAACAATGCC	tccaaAGAAG	ACCAACTCAA	ACAAGGTTAT
2251	GGGCTGAGCA	GGGTTTCCGC	CCTGCCGCGA	GATTACGGAC	GTTTGGGAAGT
2301	CGGTACGCGC	TGGTTGGGCA	ACAAACTGAC	TTTGGGCGGC	GCGATgcGCT
2351	ATTCGGCAA	GAGCATCCGC	GCGACGGCTG	AAGAACGCTA	TATCGACGGC
2401	ACCAACGGGG	GAAATACCAG	CAATGTCCGG	CAACTGGGCA	AGCGTTCCAT
2451	CAAACAAACC	GAAACCCCTG	CCCGACAGCC	TTTGATTTTT	GATTTTACG
2501	CCGCTTACGA	GCCGAAGAAA	AACCTTATTT	TCCGCGCCGA	AGTCAAAAAC
2551	CTGTTGACA	GGCGTTATAT	CGATCCGCTC	GATGCGGGCA	ATGATCGGGC
2601	AACGCAGCGT	TATTACAGCT	CGTTCGACCC	GAAAGACAAG	GACGAAGACG
2651	TACGTGTAA	TGCTGATAAA	ACGTTGTGCA	ACGGCAAATA	CGGCGGCACA
2701	AGCAAAAGCG	TATTGACCAA	TTTCGCACGC	GGACGCACCT	TCTTGATGAC
2751	GATGAGCTAC	AAGTTTTAA			

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

25
30
35
40

1	MRSSFRLKPI	CFYLMGVMLY	HHSYAEDAGR	AGSEAQIQVL	EDVHVAKAKRV
51	PKDKKVFTDA	RAVSTRQDVF	KSGENLDNIV	RSIPGAFTQQ	DKSSGIVSLN
101	IRGDSGFGRV	NTMVDGITQT	FYSTSTDAGR	AGGSSQFGAS	VDSNFIAGLD
151	VVKGSFSGSA	GINSLAGSAN	LRTLGVDDV	QGNNTYGLLL	KGLTGTNSTK
201	GNAMAAIGAR	KWLESGASVG	VLYGHSRRGV	AQNYRVGGGG	QHIGNFGEY
251	LERRKQYFV	QEGGLKFNAG	SGKWERDLQR	QYWKTKWYKK	YEDPQELQKY
301	IEEHDKSWRE	NLAPQYDITP	IDPSGLKQOS	AGNLFKLEYD	GVFNKYTAQF
351	LDLNTRIGSR	KIINRNYQFN	YGLSLNPTYN	LNLTAAYNNG	RQKYPKGAKF
401	TGWGLLKDFE	TYNNAKILD	NNTATFRLPR	ETELQTTLGF	NYFHNEYGKN
451	RFPEELGLFF	DGPDQDNGLY	SYLGRFKGDK	GLLPQKSTIV	QPAGSQYFNT
501	FYFDAALKKD	IYRLNYSTNA	INYRFGGEYT	GYGSENEFK	RAFGENSPAY
551	KEHCDPSCGL	YEPVLKKGK	KRANNHSVSI	SADFGDYFMP	FAGYSRTHRM
601	PNIQEMYFSQ	IGDSGVHTAL	KPERANTWQF	GENTYKKGLL	KQDDILGLKL
651	VGYSRIDNY	IHNVYGKWD	LNGDIPSWVG	STGLAYTIRH	RNFKDKVHHK
701	GFELELNVDY	GRFFTNLSYA	YQKSTOPTNF	SDASESPNNA	SKEDQLKQGY
751	GLSRVSALPR	DYGRLEVGT	WLGKLTLLG	AMRYFGKSIR	ATAEERYIDG
801	TNGGNTSNVR	QLGKRSIKQT	ETLARQPLIF	DFYAAYEPKK	NLI FRAEVKN
851	LFDRRYIDPL	DAGNDAATQR	YYSSFDPKDK	DEDVTCNADK	TLCNGKYGGT
901	SKSVLTNFAR	GRTFLMTMSY	KF*		

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

45
50
55
60
65

orf133ng-1.pep	10	20	30	40	50	60
orf133ng-1.pep	SFRLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLE	EDVHVAKAKRV	PKDKKVFTDARAV			
orf133-1				EAQIQVLE	EDVHVAKAKRV	PKDKKVFTDARAV
				10	20	30
orf133ng-1.pep	70	80	90	100	110	120
orf133ng-1.pep	STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS					
orf133-1	STRQDIFKSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS					
	40	50	60	70	80	90
orf133ng-1.pep	130	140	150	160	170	180
orf133ng-1.pep	TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN					
orf133-1	TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN					
	100	110	120	130	140	150
orf133ng-1.pep	190	200	210	220	230	240
orf133ng-1.pep	NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRGVAQNYRVGGGGQHI					
orf133-1	NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRGVAQNYRVGGGGQHI					
	160	170	180	190	200	210
orf133ng-1.pep	250	260	270	280	290	300
orf133ng-1.pep	GNFGEYLERRKQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWYKYEDPQELQKYIEE					

orf133-1 GNFGAEYLERRKQRYFVQEGALKFNSDSGKWERDLQRQWQKYPKYNNYNN-QELQKYIEE
 220 230 240 250 260

5 orf133ng-1.pep HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
 orf133-1 HDKSWRENLPQYDITPIDPSSLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
 270 280 290 300 310 320

10 orf133ng-1.pep NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
 orf133-1 NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDNLNT
 330 340 350 360 370 380

15 orf133ng-1.pep ATFRLPRETELQTTLGFNYFHNEYGKNRFPPEELGLFFDGPDDQDNGLYSYLGRFKGDKGLL
 orf133-1 ATFRLPRETELQTTLGFNYFHNEYGKNRFPPEELGLFFDGPDDQDNGLYSYLGRFKGDKGLL
 390 400 410 420 430 440

20 orf133ng-1.pep POKSTIVQAGSQYFNTFFYDAALKKDIYRLNYSNTNAINYRFGGEYTGYYGSENEFKRAF
 orf133-1 POKSTIVQAGSQYFNTFFYDAALKKDIYRLNYSNTNVTGYRFGGEYTGYYGSDDEFKRAF
 450 460 470 480 490 500

25 orf133ng-1.pep GENSPAYKEHCDPSCGLYEPVLKKGKRRANNHVSISADFGDYFMPFAGYSRTHRMPNI
 orf133-1 GENSPYKHKCNRSCTGIYEPVLKKGKRRANNHVSISADFGDYFMPFASYSRTHRMPNI
 510 520 530 540 550 560

30 orf133ng-1.pep QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHN
 orf133-1 QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN
 570 580 590 600 610 620

35 orf133ng-1.pep VYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
 orf133-1 VYGKWWDLNGDIPSWVSSSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
 630 640 650 660 670 680

40 orf133ng-1.pep STQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMR
 orf133-1 STQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMR
 690 700 710 720 730 740

45 orf133ng-1.pep YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAAYEPKKNLI
 orf133-1 YFGKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDYAAAYEPKKNLI
 750 760 770 780 790 800

50 orf133ng-1.pep FRAEVKNLFDRLYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGYYGTSKS
 orf133-1 FRAEVKNLFDRLYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGYYGTSKS
 810 820 830 840 850 860

55 orf133ng-1.pep VLTNFARGRTFLMTMSYKFX
 orf133-1 VLTNFARGRTFLMTMSYKFX
 870 880

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

sp|P45114|YC17_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR
>gi|1075372|piF|G64110 transferrin binding protein 1 precursor (tbpl) homolog -
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913
Score = 930 bits (2377), Expect = 0.0
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

Query: 38 QVLEDVHVKAARVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V
Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGSV 88

Query: 98 SLNIRGDSGFGRVNTMVDGITQTFYSTSTAGRAGGSSQFGASVDSNFIAGLDVVKGSFS 157
S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS
Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGGSSQFGAIDPNFIAGVDVKNKSNFS 148

Query: 158 GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESQA 217
G++GIN+LAGSAN RTLGV+DV+ M RKWL++G
Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLDNGG 208

Query: 218 SVGVLYGHSRRGVAQNYRVGGGQHGIFGEEYLERRKQYFVQEGGLKFNAGSGKWERD 277
VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D
Sbjct: 209 YGVVYGYSGREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTPD 265

Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303
L +++W +Y KK +D ++LQK IEE
Sbjct: 266 LSKKHWSCNKPQYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDIEE 325

Query: 304 HDKSWRENLAQYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI
Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEGDDHQNLAQLRTLONKIGSRKIE 384

Query: 364 NRNYQFNYSGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT 423
NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+
Sbjct: 385 NRNYQVNYNFNNSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

Query: 424 ATFRLPRETELQTLGFFNYFHNHYGKGRFPEELGLFFDGDQDNGLYSY--LGRFGKDGK 481
TF LP+E +L+TTLGFNYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K
Sbjct: 445 HTFLLPKEIDLKTTLGFNYFTNEYSKNRFPEELSLFYNDASHDQGLYSHSKRGRYSGTGS 504

Query: 482 LLPQKSTIVQPAQSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR 541
LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY
Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDTALSKGIYHLNYSVNFTHYAFNGEYVGY----- 555

Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKYKGRANNHVSISADFGDYFMPFAGYSRTHRMP 601
EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP
Sbjct: 556 ---ENTAGQQ-----INEPILHKS GHKKA FNHSATLSAELSDYFMPFFTYSRTHRMP 604

Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661
NIQEM+FSQ+ ++GV+TALKPE++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI
Sbjct: 605 NIQEMFFSQVSNAQVNTALKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNI 664

Query: 662 HNVYKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 721
HNVYK WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY
Sbjct: 665 HNVYGVWW--RDGMPTWAESNGFKYTIHQYKPIVKKSGVELEINYDMGRFFANVSAY 722

Query: 722 QKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGRWLGNKLTGLGA 781
Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTGL A
Sbjct: 723 QRTNQPTNYADASPRNNASQEDILKQGYGLSRVSMPLPKDYGRLELGRWFDQKLTGLA 782

Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAAYEPKKN 841
RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+
Sbjct: 783 ARYYGKSKRATIEEYINGSR-FKKNLRLRENNYAVKKTEDIKKQPIILDLHVSYPEIKD 841

Query: 842 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGKYGGS 901
LI +AEV+NL D+RY+DPLDAGNDAA+QRYYS + + C D + C GG+
Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYYSL-----NNSIECAQDSSAC----GGSD 892

Query: 902 KSVLTNFAARGRTFLMTMSYKF 922
K+VL NFARGRT+++++YKF
Sbjct: 893 KTVLYNFAARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1  ATGAACCTGA  TTTCACGTTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
101 ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAAATGCTG
151 GGCTACACCG  CCCTCAAAAT  GCCCGCCCGC  GCCTACGAAC  TGATTCCCCT
201 CGCCGTCCTT  ATCGGCGGAC  TGGTCTCCCT  CAGCCAGCTT  GCCGCCGGCA
251 GCGAACTGAC  CGTCATCAAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301 TTGATTCTGT  CGCAGTTCGG  TTTTATTTT  GCTATTGCCA  CCGTCGCGCT
351 CGGCGAATGG  GTTGCGCCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
401 CCGCCGCCAT  CAACGGCAAA  ATCAGCACCG  GCAATACCGG  CCTTTGGCTG
451 AAAGAAAAAA  ACAGCGTGAT  CAATGTGCGC  GAAATGTTGC  CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1  MNLISRYIIR  QMAVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
51  GYTALKMPAR  AYELIPLAVL  IGGLVLSLSQL  AAGSELTVIK  ASGMSTKKLL
101 LILSQFGFIF  AIATVALGEW  VAPTLSQKAE  NIKAAAINGK  ISTGNTGLWL
151 KEKNSVINVR  EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1  ATGAACCTGA  TTTCACGTTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
101 ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAAATGCTG
151 gGCTACACCG  CCCTCAAAAT  GCCCGCCCGC  GCCTACGAAC  TGATTCCCCT
201 CGCCGTCCTT  ATCGGCGGAC  TGGTCTCCCT  CAGCCAGCTT  GCCGCCGGCA
251 GCGAACTGAC  CGTCATCAAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301 TTGATTCTGT  CGCAGTTCGG  TTTTATTTT  GCTATTGCCA  CCGTCGCGCT
351 CGGCGAATGG  GTTGCGCCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
401 CCGCCGCCAT  CAACGGCAAA  ATCAGCACCG  GCAATACCGG  CCTTTGGCTG
451 AAAGAAAAAA  ACAGC rTKAT  CAATGTGCGC  GAAATGTTGC  CCGACCATAC
501 GCTTTTGGGC  ATCAAAATTT  GGGCGCGCAA  CGATAAAAC  GAATTGGCAG
551 AGGCAGTGGA  AGCCGATTCC  GCCGTTTTGA  ACAGCGACGG  CAGTTGGCAG
601 TTGAAAAACA  TCCGCCGCAG  CACGCTTGCC  GAAGACAAAG  TCGAGGTCTC
651 TATTGCGGCT  GAAGAAAACT  GGCCGATTTC  CGTCAAACGC  AACCTGATGG
701 ACGTATTGCT  CGTCAAACCC  GACCAAATGT  CCGTCGGCGA  ACTGACCACC
751 TACATCCGCC  ACCTCCAAAA  CAACAGCCAA  AACACCCGAA  TCTACGCCAT
801 CGCATGGTGG  CGCAAATTGG  TTTACCCCGC  CGCAGCCTGG  GTGATGGCGC
851 TCGTCGCCTT  TGCCTTTACC  CCGCAAACCA  CCCGCCACGG  CAATATGGGC
901 TTAATACTCT  TCGGCGGCAT  CTGTsTCGGA  TTGCTGTTCC  ACCTTGCCGG
951 ACGGCTCTTT  GGGTTTACCA  GCCAACTCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1  MNLISRYIIR  QMAVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
51  GYTALKMPAR  AYELIPLAVL  IGGLVLSLSQL  AAGSELTVIK  ASGMSTKKLL
101 LILSQFGFIF  AIATVALGEW  VAPTLSQKAE  NIKAAAINGK  ISTGNTGLWL
151 KEKNSXINVR  EMLPDHTLLG  IKIWARNDKN  ELAEAVEADS  AVLNSDGSWQ
201 LKNIRRTSLG  EDKVEVSIAA  EENWPISVKR  NLMDVLLVKP  DQMSVGEALT
251 YIRHLQNSQ  NTRIIAIAWW  RKLVPAAAW  VMALVAEAF  POTTRHGNMG
301 LKLFGGICXG  LLFHLAARLF  GFTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.*

meningitidis:

5	orf112.pep	10	20	30	40	50	60
	orf112a	10	20	30	40	50	60
10	orf112.pep	70	80	90	100	110	120
	orf112a	70	80	90	100	110	120
15	orf112.pep	130	140	150	160		
	orf112a	130	140	150	160		
20	orf112a	170	180				
	orf112a	190	200	210	220	230	240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTT	GAAATCCTGT
30	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAAC	TGATGCCCCCT
35	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGAN	CGTCATCAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
40	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTT	GCTATTGCCA	CCGTCGCGCT
	351	CGGCGAATGG	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
45	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
50	501	CCTGCTGGGC	ATTAATCT	GGGCCCCGAA	CGATAAAAC	GAAGTGGCAG
	551	AGGCAAGTGA	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	CAGTTGGCAG
55	601	TTGAAAAACA	TCCGCCGCG	CACGCTTGGC	GAAGACAAAG	TCGAGGTCTC
	651	TATTGCGGCT	GAAGAAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
60	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTCGGCGA	ACTGACCACC
	751	TACATCCGCC	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	TCTACGCCAT
65	801	CGCATGGTGG	CGCAAATTGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCGTCGCCTT	TGCCTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
70	901	TTAAAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCTCTCG
75	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

50	1	MNLISRYIIR	QMAVMAVYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMXAR	AYELMPLAVL	IGGLVSXSQ	AAGSELXVIK	ASGMSTKKLL
55	101	LILSQFGFIF	AIATVALGEW	VAPTLSQKAE	NIKAAAINGK	ISTGNTGLWL
	151	KEKNSIINVR	EMPLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSWQ
60	201	LKNIRRTL	EDKVEVSIAA	EEXWPISVKR	NLMDVLLVKP	DQMSVGELTT
	251	YIRHLQXXSQ	NTRIYAIWW	RKLVPAAAW	VMALVAFAT	PQTRRHGNGM
65	301	LKXFGGICLG	LLFHLAGRLF	XFTSQLYGIP	PFLXGALPTI	AFALLAVWLI
	351	RKQEKRT				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

60	orf112a.pep	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
	orf112-1	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
65	orf112a.pep	AYELMPLAVLIGGLVSXSQLAAGSELXVIKASGMSTKKLLILSQFGFIFAIATVALGEW
	orf112-1	AYELMPLAVLIGGLVSXSQLAAGSELXVIKASGMSTKKLLILSQFGFIFAIATVALGEW

5

10

15

20

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N.*

gonorrhoeae:

25

30

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

35

40

45

50

55

This encodes a protein having amino acid sequence <SEQ ID 892>:

60

-486-

351 RKQEK R*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLISRYIIQMAVMVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
	orf112-1	MNLISRYIIQMAVMVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
		10	20	30	40	50	60
10	orf112ng	70	80	90	100	110	120
	orf112-1	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLLLSQFGFIFAIAAVALGEW					
		70	80	90	100	110	120
15	orf112ng	130	140	150	160	170	180
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN					
		130	140	150	160	170	180
20	orf112ng	190	200	210	220	230	240
	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRSIMGTDKIETSAAAEETWPIAVRRNLMVLLVKP					
25		190	200	210	220	230	240
	orf112ng	250	260	270	280	290	300
	orf112-1	DQMSVGELTTYIRHLQNNNQNTQIYAIWWRKLVYPVAAWVMALVAFATPQTTRHGNMG					
30		250	260	270	280	290	300
	orf112ng	310	320	330	340	350	
35	orf112-1	LKLFGGICLGLLFHLAGRLFGFTSOLYGTTPFLAGALPTIAFALLAVWLIRKQEKRX					
		310	320				

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG	BamHI-NheI XhoI
	Reverse	CCC <u>GCTCGAG</u> -CCAGCGGTAGCCTAATT	
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -GACGGCATAACGGCG	
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -TGATTTACGGACGCGCA	
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -TTTGGCTGCGCCTTC	
ORF 5	Forward	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAAACC	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAAC	
	Reverse	CCC <u>GCTCGAG</u> -GACTGTGCAAAAACGG	
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -TGCGCCGAACACTTTC	
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTTTGTTC	BamHI-NheI XhoI
	Reverse	CCC <u>GCTCGAG</u> -TTTCAAAATATATTTGCGGA	
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -AGCAGGCTTTGGCGC	
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -TTTCCGAGGTTTTTCGGG	
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -TAATGGGAAACCTTGTTTT	
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -GGAAACGACTTCGCC	
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC	BamHI-NdeI XhoI
	Reverse	CCC <u>GCTCGAG</u> -AGGGTGTGTGATAATAAG	
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG	
	Reverse	CCC <u>GCTCGAG</u> -AGGTTGGCCTTGCTATG	
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCCGGCCTGTTTCG	NdeI-NcoI

	Forward Reverse	C <u>GGGATCC</u> -ATTGCCGGCCTGTTCCG CCC <u>GCTCGAG</u> -AAGCAGGTTGTACAGC	BamHI XhoI
ORF 18	Forward Reverse	G <u>CGGATCCCATATG</u> -ATTTTGCTGCATTTGGAT CCC <u>GCTCGAG</u> -TCTTCCAATTTCTGAAAGC	BamHI-NdeI XhoI
ORF 19	Forward Forward Reverse	GGAATTCATATGGCCATGG -TCGCCAGTGTTTTTACC C <u>GGGATCC</u> -TTCGCCAGTGTTTTTACCG CCC <u>GCTCGAG</u> -GGTGTGTTTTGAAGCTGCC	NdeI-NcoI BamHI XhoI
ORF 20	Forward Forward Reverse	GGAATTCATATGGCCATGG -TCGGCGCGGGTATG C <u>GGGATCC</u> -TTCGGCGCGGGTATG CCC <u>GCTCGAG</u> -CGGCGAGCGAGAGCA	NdeI-NcoI BamHI XhoI
ORF 22	Forward Forward Reverse	GGAATTCATATGGCCATGG-TGATTAAATCAAAAAGGTCT C <u>GGGATCC</u> -ATGATTAAATCAAAAAGGTCTAAACC CCC <u>GCTCGAG</u> -ATTATGATAGCGGCC	NdeI-NcoI BamHI XhoI
ORF 23	Forward Reverse	C <u>GCGGATCCCATATG</u> -GATGTTTCTGTTTCAGAC CCC <u>GCTCGAG</u> -TTTAAACCGATAGGTAAACG	BamHI-NdeI XhoI
ORF 24	Forward Forward Reverse	GGAATTCATATGGCCATGG -TGATGCCGGAATGGTG C <u>GGGATCC</u> -ATGATGCCGGAATGGTG CCC <u>GCTCGAG</u> -TGTCAGCGTGGCGCA	NdeI-NcoI BamHI XhoI
ORF 25	Forward Reverse	G <u>CGGATCCCATATG</u> -TATCGCAAATGATTGC CCC <u>GCTCGAG</u> -ATCGATGGAATAGCCG	BamHI-NdeI XhoI
ORF 26	Forward Reverse	G <u>CGGATCCCATATG</u> -CAGCTGATCGACTATTC CCC <u>GCTCGAG</u> -GACATCGGCGCGTTTT	BamHI-NdeI XhoI
ORF 27	Forward Forward Reverse	GGAATTCATATGGCCATGG-AGACCTATTCTGTTTA C <u>GGGATCC</u> - CAGACCTATTCTGTTTATTTAATC CCC <u>GCTCGAG</u> -GGGTTCGATTAAATAACCAT	NdeI-NcoI BamHI XhoI
ORF 28	Forward Forward Reverse	GGAATTCATATGGCCATGG-ACGGCTGTACGTTGATGT C <u>GGGATCC</u> -AACGGCTGTACGTTGATG CCC <u>GCTCGAG</u> -TTTGTCTAGAGGAATTCGCG	NdeI-NcoI BamHI XhoI
ORF 29	Forward Forward Reverse	G <u>CGGATCCCATATG</u> -AACGGTTTGGATGCCCC C <u>GCGGATCCGCTAGC</u> -AACGGTTTGGATGCCCC CCC <u>GCTCGAG</u> -TTTGTCTAAGTTCCTGATATG	BamHI-NdeI BamHI-NheI XhoI
ORF 32	Forward Reverse	C <u>GCGGATCCCATATG</u> -AATACTCCTCCTTTTG CCC <u>GCTCGAG</u> -GCGTATTTTTTGATGCTTTG	BamHI-NdeI XhoI
ORF 33	Forward Reverse	G <u>CGGATCCCATATG</u> -ATTGATAGGGATCGTATG CCC <u>GCTCGAG</u> -TTGATCTTTCAAACGGCC	BamHI-NdeI XhoI

ORF 35	Forward Forward Reverse	GCGGATCCCATATG-TTCAGAGCTCAGCTT CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT CCCGCTCGAG-AAACAGCCATTTGAGCGA	BamHI-NdeI BamHI-NheI XhoI
ORF 37	Forward Reverse	GCGGATCCCATATG-GATGACGTATCGGATTTT CCCGCTCGAG-ATAGCCCGCTTTCAGG	BamHI-NdeI XhoI
ORF 58	Forward Reverse	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT CCCGCTCGAG-AGCATTGTCCAAGGGGAC	BamHI-NheI XhoI
ORF 65	Forward Forward Reverse	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG CGGGATCC-TTGCTGTATCTGAATCAAGG CCCGCTCGAG-CCGCATCGGCAGACA	 NdeI-NcoI BamHI XhoI
ORF 66	Forward Reverse	GCGGATCCCATATG-TACGCATTTACCGCCG CCCGCTCGAG-TGGATTTTCAGAGATGG	BamHI-NdeI XhoI
ORF 72	Forward Reverse	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA CCCGCTCGAG-GCCTGAGACCTTTGCAA	BamHI-NdeI XhoI
ORF 73	Forward Reverse	GCGGATCCCATATG-AGATTTTTCGGTATCGG CCCGCTCGAG-TTCATCTTTTTCATGTTTCG	BamHI-NdeI XhoI
ORF 75	Forward Reverse	GCGGATCCCATATG- TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTTGAAGACAG	BamHI-NdeI XhoI
ORF 76	Forward Reverse	GATCAGCTAGCCATATG-AAACAGAAAAAAACCGC CGGGATCC-TTACGGTTTGACACCGTT	NheI-NdeI BamHI
ORF 79	Forward Reverse	CGCGGATCCCATATG-GTTTCCGCCGCCG CCCGCTCGAG-GTGCTGATCGGCTTCG	BamHI-NdeI XhoI
ORF 83	Forward Reverse	GCGGATCCCATATG-AAAACCCTGCTGCTGC CCCGCTCGAG-GCCGCCTTTGCGGC	BamHI-NdeI XhoI
ORF 84	Forward Reverse	GCGGATCCCATATG-GCAGAGATCTGTTTG CCCGCTCGAG-GTTTGCCGATCCGACCA	BamHI-NdeI XhoI
ORF 85	Forward Reverse	CGCGGATCCCATATG- GCGGTTTGGGGCGGA CCCGCTCGAG-TCGGCGCGGCGGGC	BamHI-NdeI XhoI
ORF 89	Forward Forward Reverse	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA CGGGATCC-GCCATACCTTCTTATCAGAG CCCGCTCGAG-TTTTTTGCGATTAGAAAAGC	NdeI-NcoI BamHI XhoI
ORF 97	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

ORF 98	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
	Forward Reverse	GCGGATCCCATATG-ACGGTAACTGCGG CCCGCTCGAG-TTGTTGTTGCGGGCAAATC	BamHI-NdeI XhoI
ORF 100	Forward Reverse	GCGGATCCCATATG-TCGGGCATTTACACCG CCCGCTCGAG-ACGGGTTTCGGCGGAA	BamHI-NdeI XhoI
ORF 101	Forward Reverse	GCGGATCCCATATG-ATTTATCAAAGAAACCTC CCCGCTCGAG-TTTTCGCCTTTCAATGT	BamHI-NdeI XhoI
ORF 102	Forward Reverse	GCGGATCCCATATG-GCAGGGCTGTTTTACC CCCGCTCGAG-AAACGGTTTGAACACGAC	BamHI-NdeI XhoI
ORF 103	Forward Reverse	GCGGATCCCATATG-AACCACGACATCAC CCCGCTCGAG-CAGCCACAGGACGGC	BamHI-NdeI XhoI
ORF 104	Forward Reverse	GCGGATCCCATATG-ACGTGGGGAACGC CCCGCTCGAG-GCGGCGTTTGAACGGC	BamHI-NdeI XhoI
ORF 105	Forward Reverse	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC CCCGCTCGAG-TAAACGAATGCCGTCCAG	BamHI-NdeI XhoI
ORF 106	Forward Reverse	GCGGATCCCATATG-AGGATAACCGACGGCG CCCGCTCGAG-TTGTTCCCGATGATGTT	BamHI-NdeI XhoI
ORF 109	Forward Reverse	GCGGATCCCATATG-GAAGATTTATATATAATACTCG CCCGCTCGAG-ATCAGCTTCGAACCGAAG	BamHI-NdeI XhoI
ORF110	Forward Reverse	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC AAACTGCAG-GGAAAACCATCCGCACTCTGCC	EcoRI PstI
ORF111	Forward Reverse	AAAGAATTC-GCACCGCAAAAGGCAAAAACCGCA AAACTGCAG-TCTGCGCGT TTTGCGGCAGGGTGG	EcoRI PstI
ORF113	Forward Reverse	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG	EcoRI PstI
ORF115	Forward Reverse	AAAGAATTC-TTGCTTGTGCAAAACAGAAAAAGACGG AAAAAGTCGAC-CTATTTTTAGGGGC TTTGCTTGTTTGAAAAGCCTGCC	EcoRI SalI
ORF119	Forward Reverse	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC	EcoRI PstI
ORF120	Forward Reverse	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG AAACTGCAG-CGGTTTGCTGCCTGGCCGTTGAT	EcoRI PstI
ORF121	Forward Reverse	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC	EcoRI PstI

ORF122	Forward Reverse	AAAAAAGTCGAC-ATGTC TTACCGCGCAAGCAGTTC TCC AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC	Sall PstI
ORF125	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG	EcoRI PstI
ORF126	Forward Reverse	AAAGAATTC-GCGGAAACGGTGAAG AAACTGCAG-TTAATCTTGTCTTCCGATATAC	EcoRI PstI
ORF127	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG AAAAAAGTCGAC-CTTAAGTAACTTGCACTCCTTATC	EcoRI Sall
ORF128	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTGAGCAGGCG	EcoRI PstI
ORF129	Forward Reverse	AAAGAATTC-ATGGATTTTCTGTTTTGACATTATTTACGAATACCG AAACTGCAG-TTATTTTTTGATGAAATTTTGGGGCGG	EcoRI PstI
ORF130	Forward Reverse	AAAGAATTC-GCAGTACTTGCCAT TCTCGGTGCG AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
ORF 131	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
ORF 132	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
ORF 133	Forward Reverse	CGCGGATCCCATATG-GAAGATGCAGGGCGCG CCCGCTCGAG-AAACTTGTAAGTCATCGT	BamHI-NdeI XhoI
ORF 134	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
ORF 135	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
ORF 136	Forward Reverse	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC CCCGCTCGAG-TTCCGAATATTTGGAACTTT	BamHI-NheI XhoI
ORF 137	Forward Reverse	CGCGGATCCCATATG-GGCACGGCGGGAAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
ORF 138	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTTATAGCGG	BamHI-NdeI XhoI
ORF 139	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

ORF 140	Forward Reverse	GCGGATCCCATATG-TTGCCACAGGCAGC CCCGCTCGAG-GACGATGGCAAACAGC	BamHI-NdeI XhoI
ORF 141	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTTAAATATT	BamHI-NdeI XhoI
ORF 142	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
ORF 143	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTCCGCAATATG	BamHI-NdeI XhoI
ORF 144	Forward Reverse	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
ORF 147	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTGTTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SaI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SaI* site was used in the reverse primer.

5

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 10 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 15 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 20 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 25 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
- 30 5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

7. An antibody which binds to a protein according to any one of claims 4 to 6.

8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.

9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
& 891..

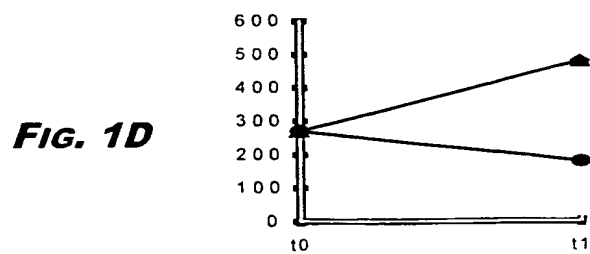
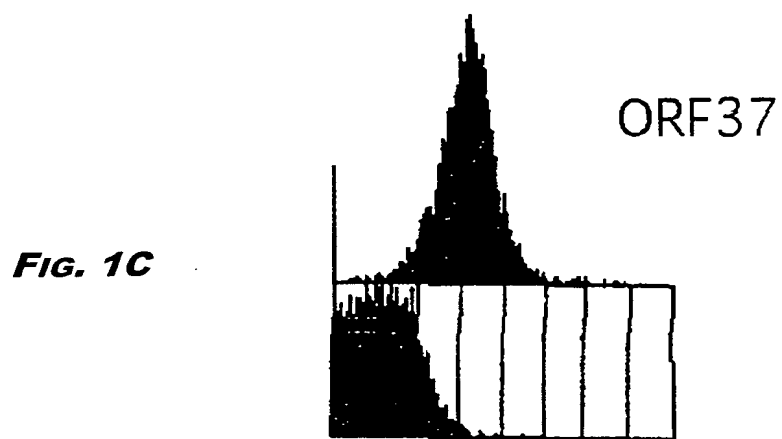
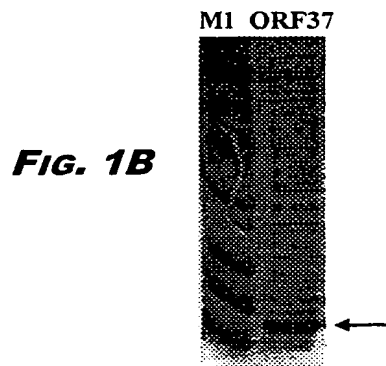
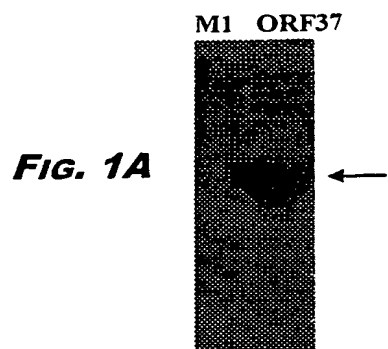
10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

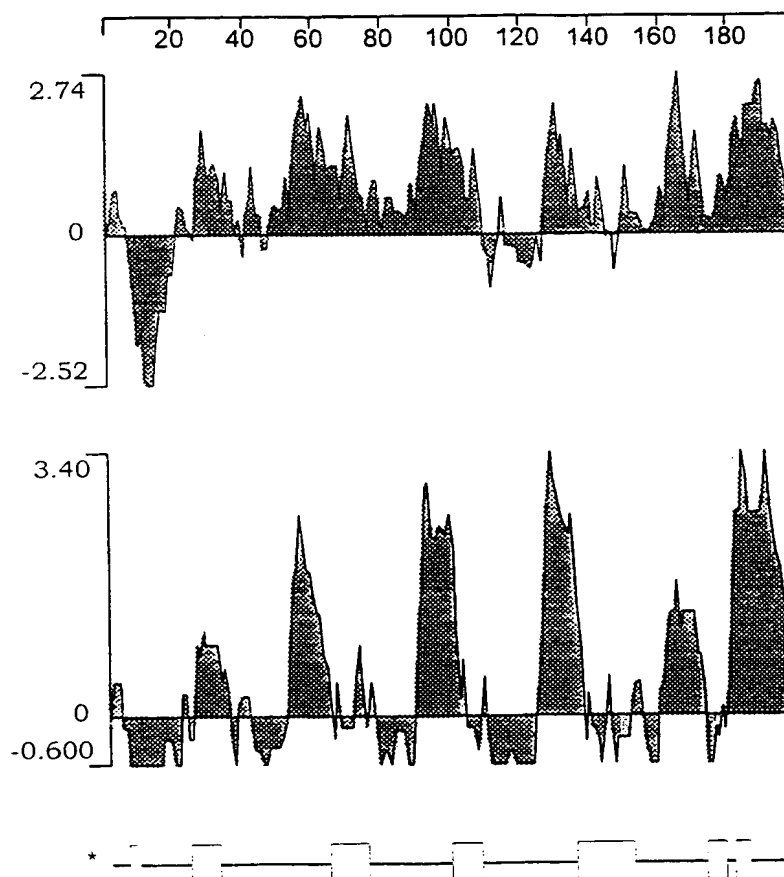
1/24

FIGURE 1



2/24

FIG 1E



3/24

FIGURE 2

FIG. 2A

M1 ORF5

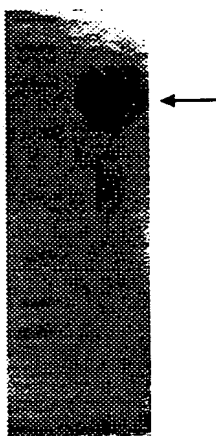
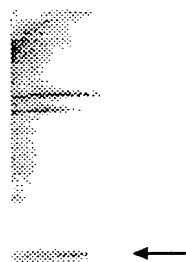


FIG. 2B

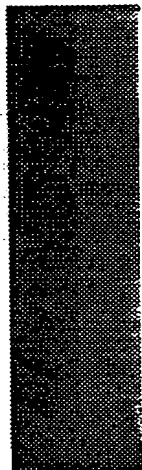
TP



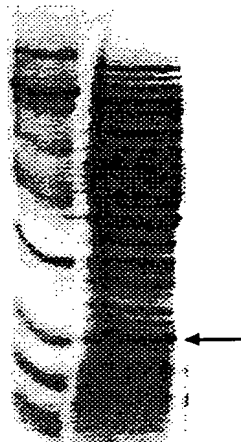
4/24

FIGURE 3**FIG. 3A**

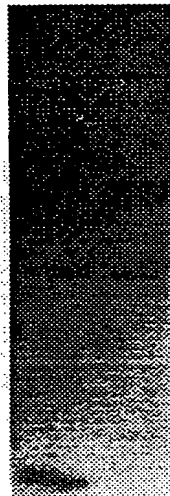
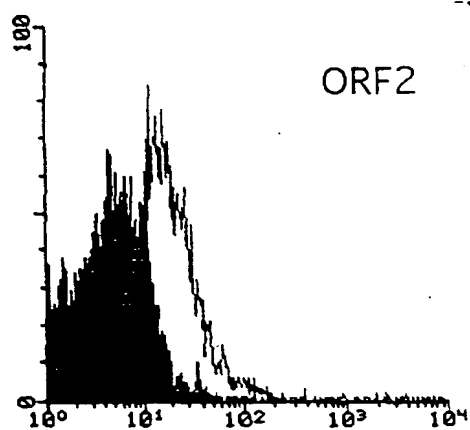
M1 ORF2

**FIG. 3B**

M1 ORF2

**FIG. 3C**

TP OMV

**FIG. 3D**

5/24

FIGURE 4

FIG. 4A

M1 ORF15

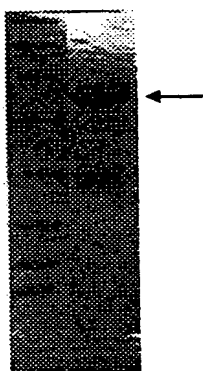


FIG. 4B

M2 ORF15

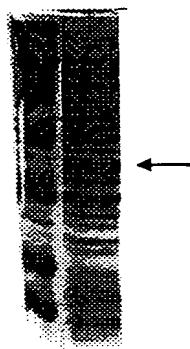
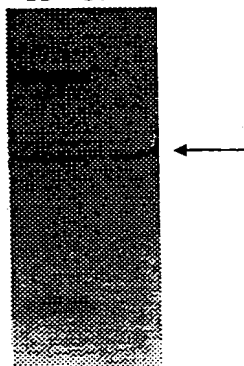


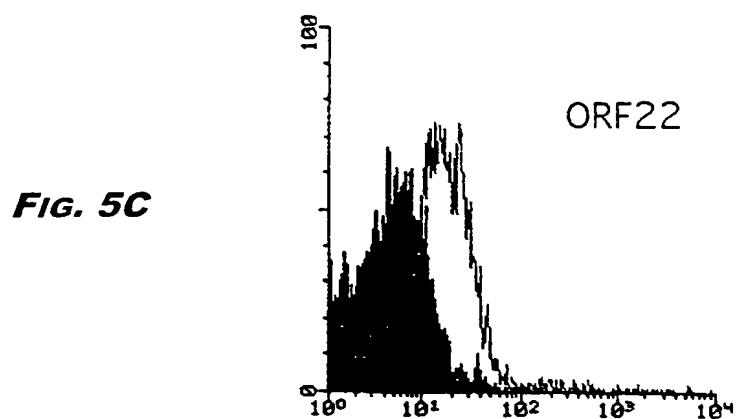
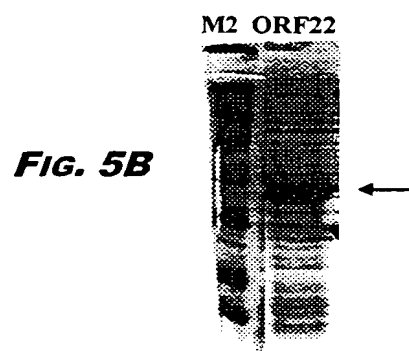
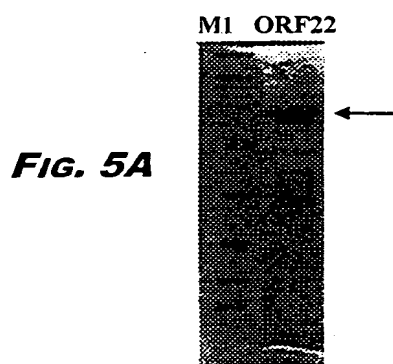
FIG 4C

TP OMV



6/24

FIGURE 5



7/24

FIGURE 6

Fig. 6A

M1 ORF28

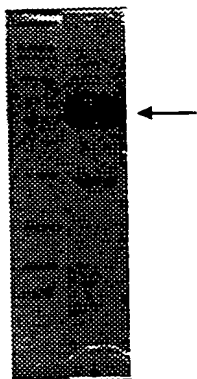


Fig. 6B

M2 ORF28

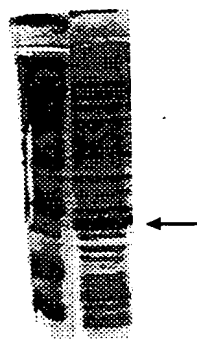


FIGURE 7

Fig. 7A

M1 ORF32

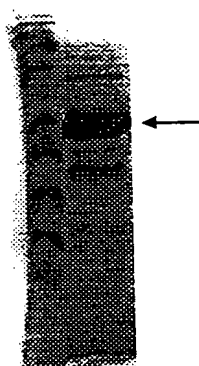
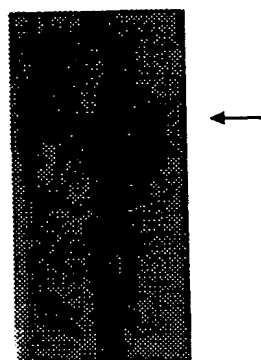


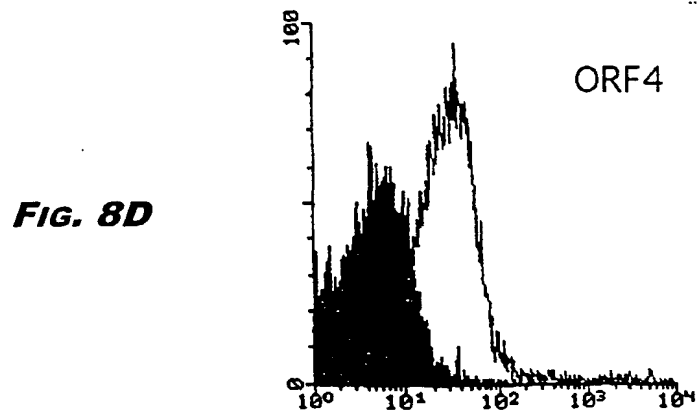
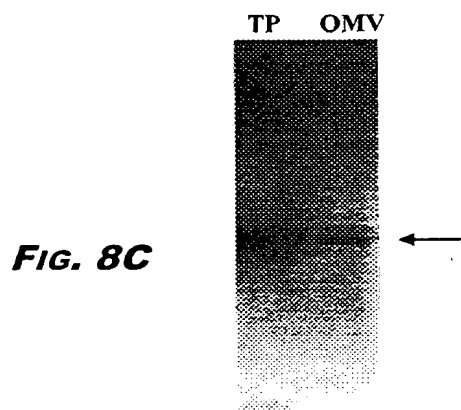
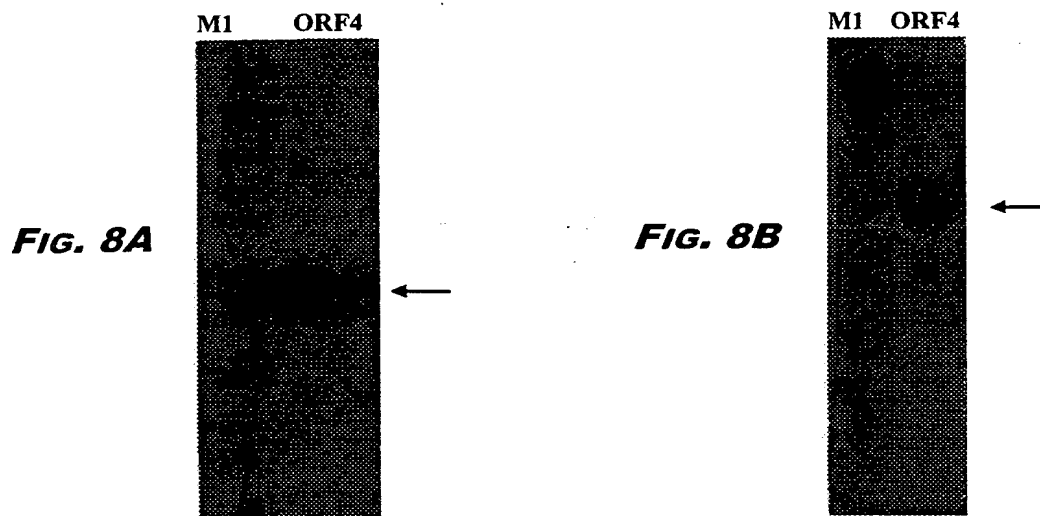
Fig. 7B

M1 ORF32

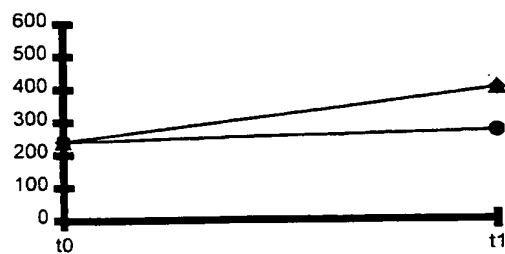
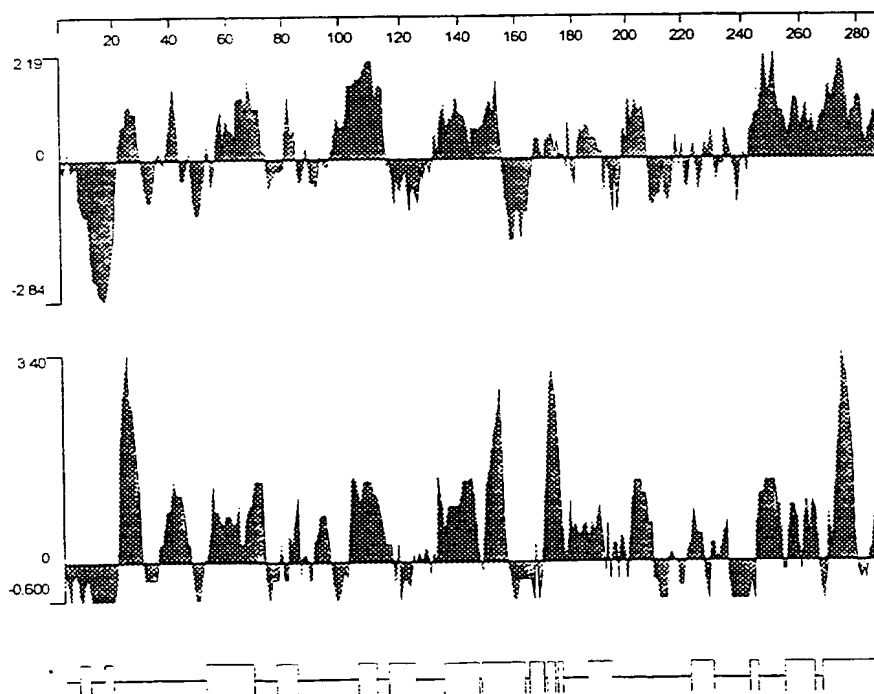


8/24 .

FIGURE 8

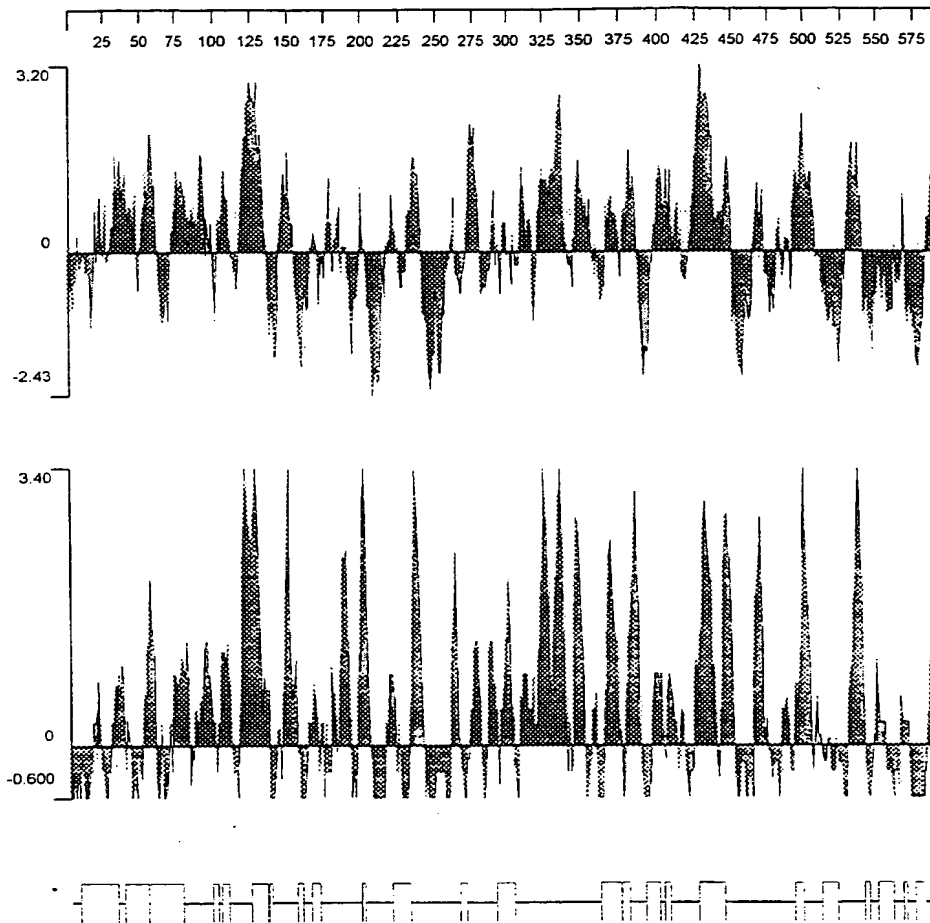


9/24

FIG. 8E**FIG. 8F**

10/24

FIGURE 9



11/24

FIGURE 10

FIG. 10A

M1 ORF76

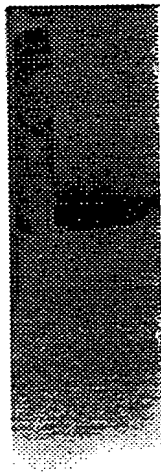


FIG. 10B

TP OMV

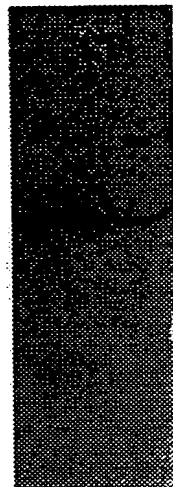
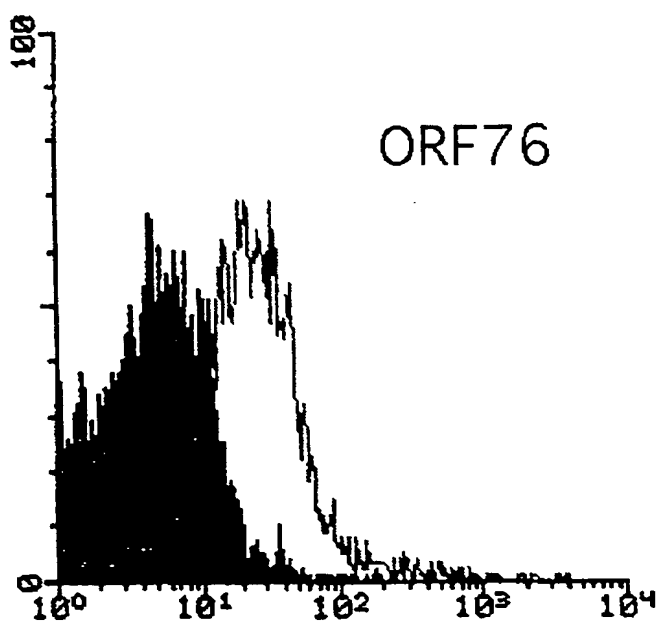


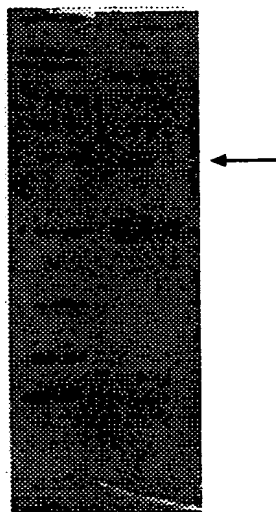
FIG. 10C



12/24

FIGURE 11

M1 ORF89



13/24

FIGURE 12

FIG. 12A

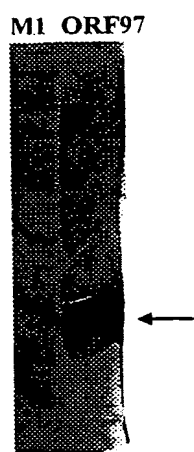


FIG. 12B

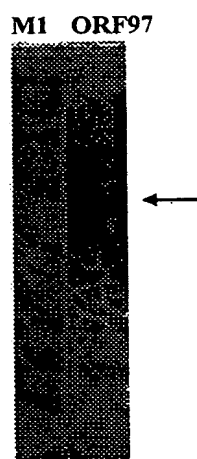


FIG. 12C

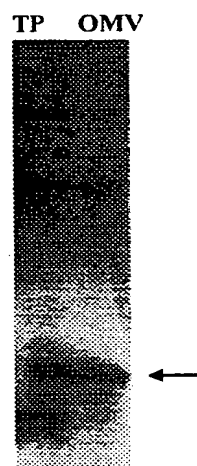
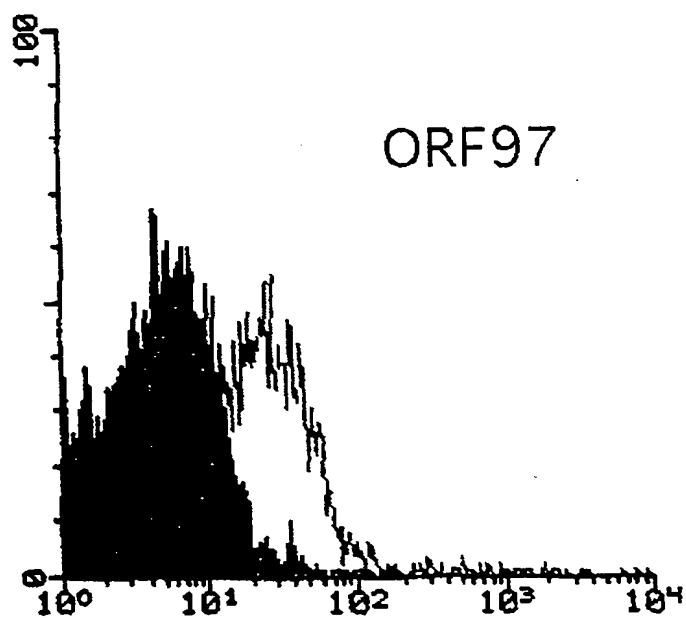
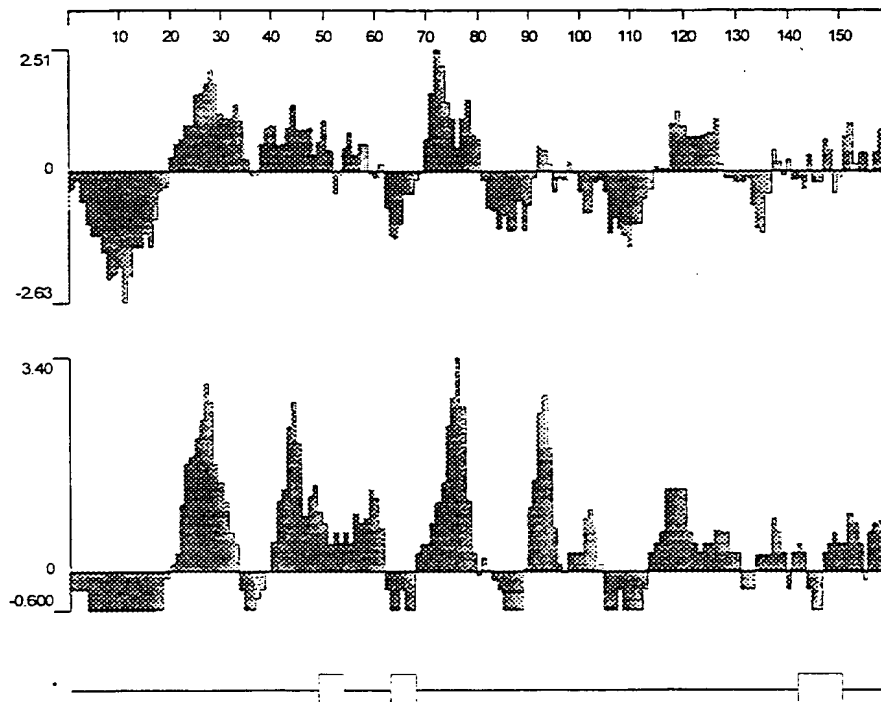


FIG. 12D



14/24

FIG. 12E

15/24

FIGURE 13

Fig. 13A

M1 ORF106

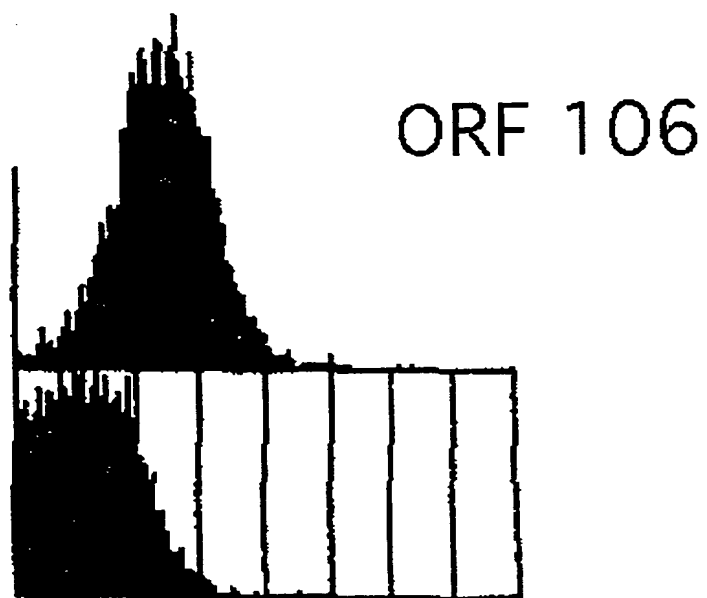


Fig. 13B

M2 ORF106



Fig. 13C



16/24

FIGURE 14

FIG. 14A

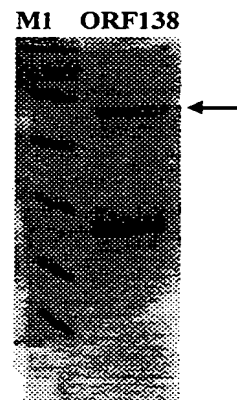
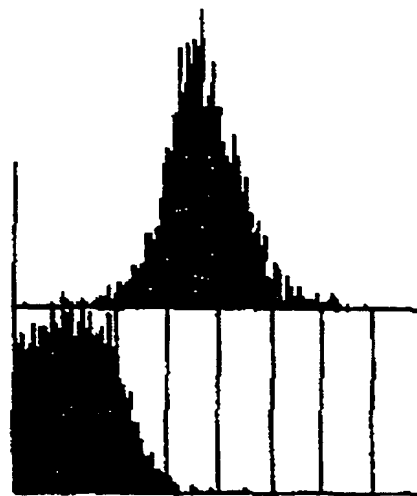


FIG. 14B



17/24

FIGURE 15

Fig. 15A

M1 ORF23

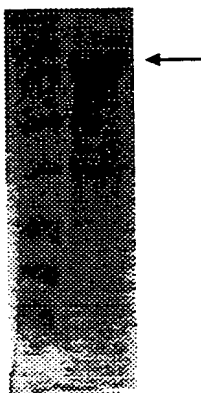


Fig. 15B

M2 ORF23

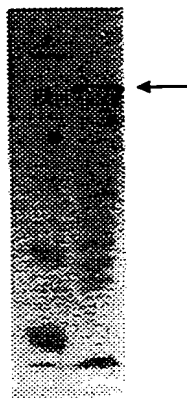
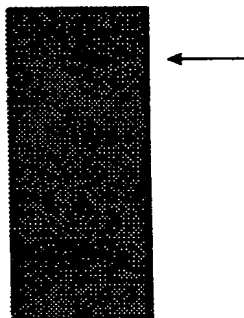


Fig 15C

TP OMV



18/24

FIGURE 16

FIG. 16A

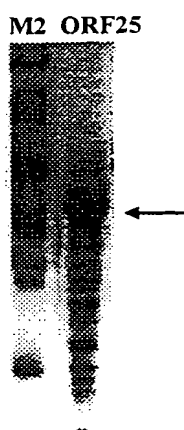


FIG. 16B

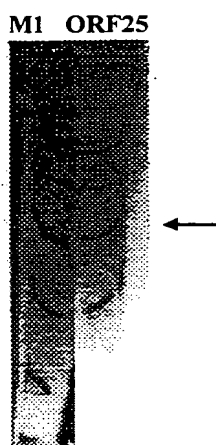


FIG. 16C

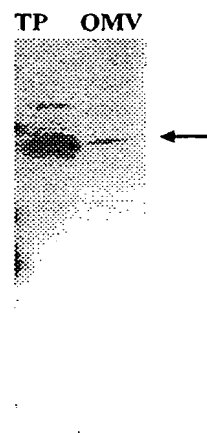
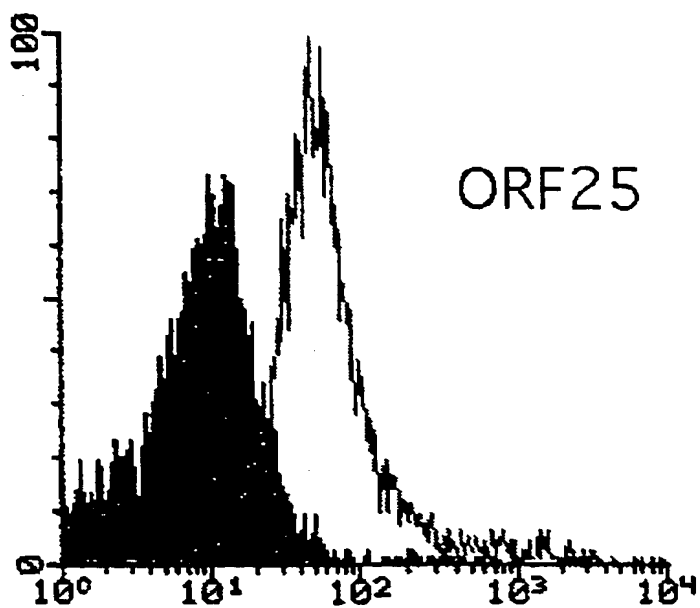
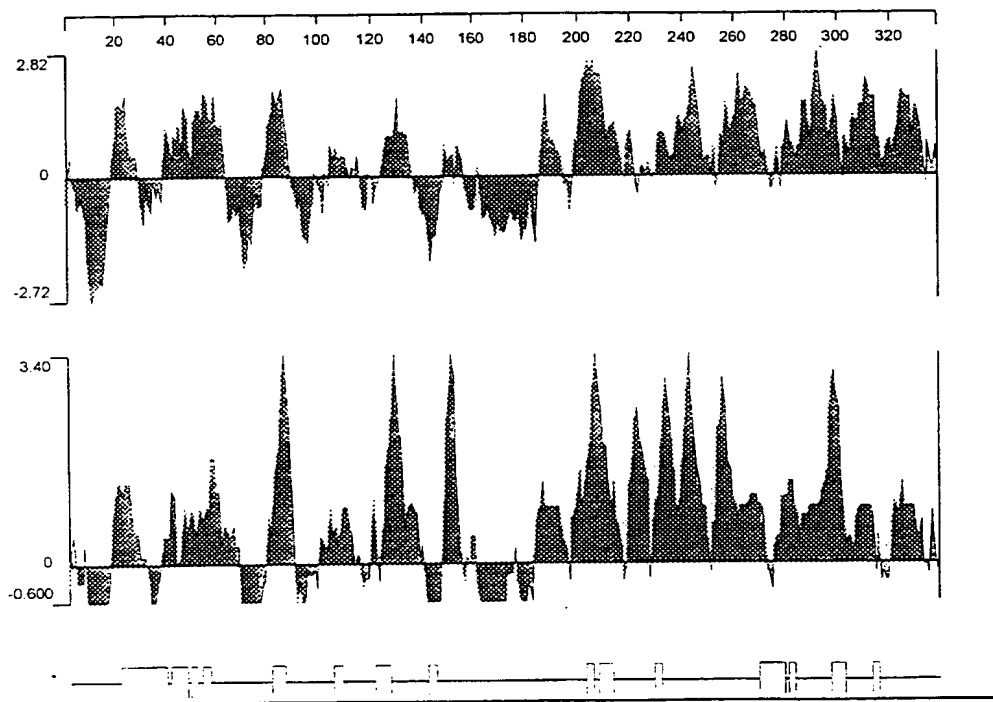


FIG. 16D



19/24

FIG. 16E

20/24

FIGURE 17

FIG. 17A

M1 ORF27

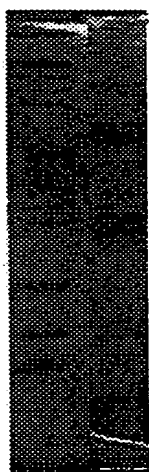


FIG. 17B

M2 ORF27



21/24

FIGURE 18

FIG. 18A

M1 ORF79

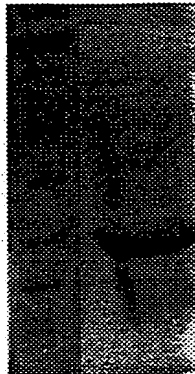
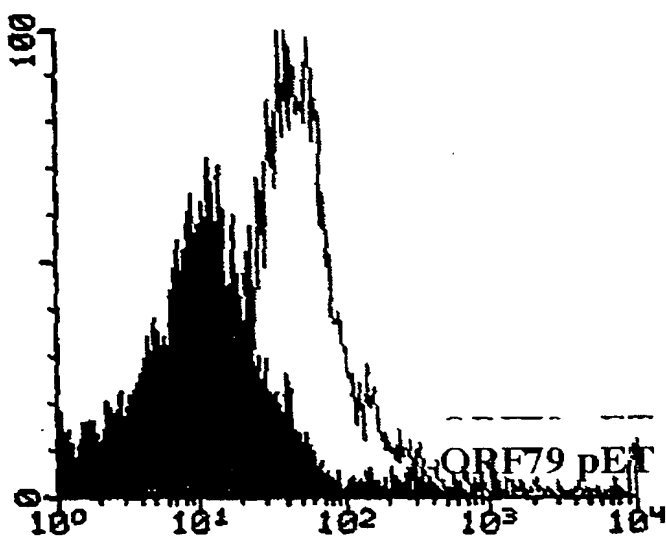


FIG. 18B



22/24

FIGURE 19

FIG. 19A

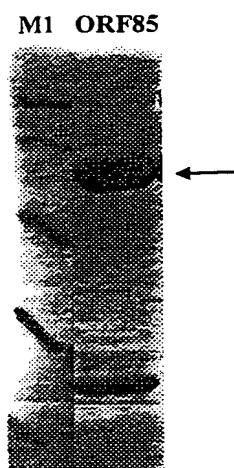


FIG. 19B

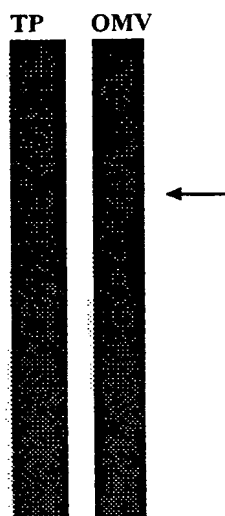
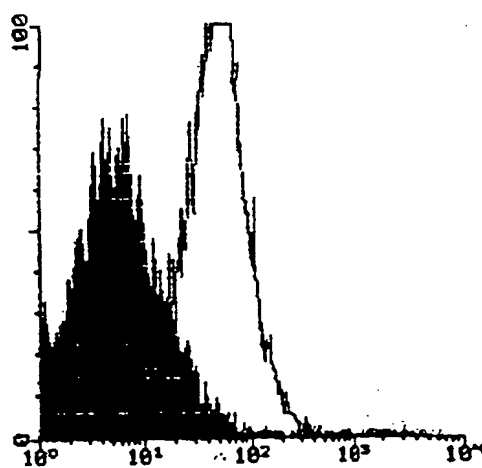
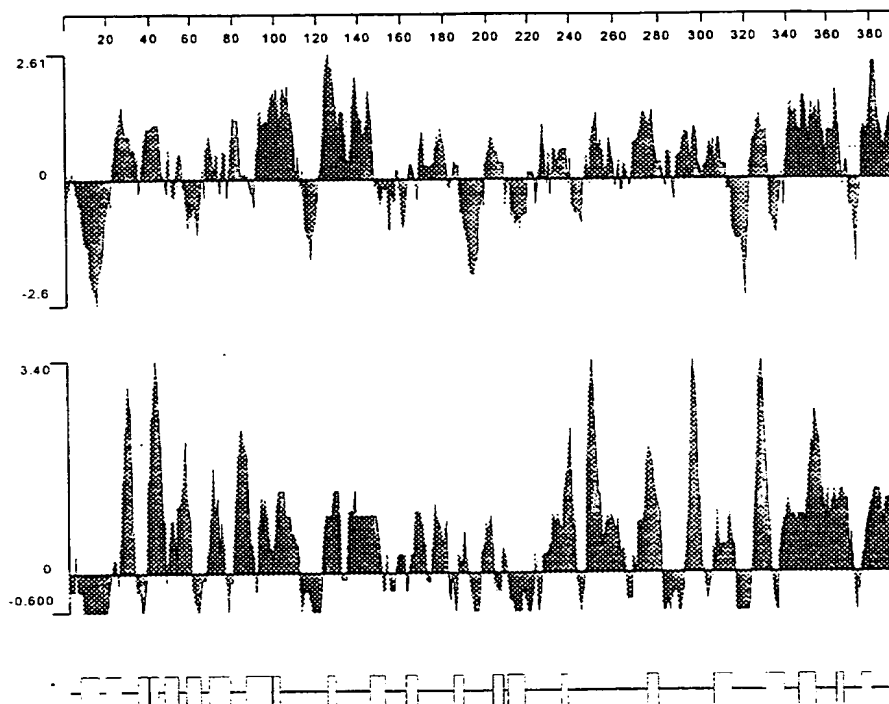


FIG. 19C



23/24

FIG 19D



24/24

FIGURE 20

Fig. 20A

M1 ORF132

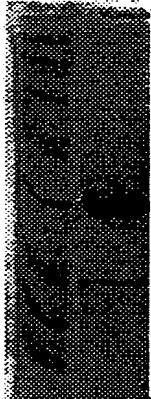


Fig. 20B

M2 ORF132



Fig. 20C





INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/22, 16/12, G01N 33/53, A61K 39/095 // (C12N 15/31, C12 R 1:36)		A3	(11) International Publication Number: WO 99/24578 (43) International Publication Date: 20 May 1999 (20.05.99)																					
(21) International Application Number: PCT/IB98/01665 (22) International Filing Date: 9 October 1998 (09.10.98)		(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).																						
(30) Priority Data: <table border="0"> <tr> <td>9723516.2</td> <td>6 November 1997 (06.11.97)</td> <td>GB</td> </tr> <tr> <td>9724190.5</td> <td>14 November 1997 (14.11.97)</td> <td>GB</td> </tr> <tr> <td>9724386.9</td> <td>18 November 1997 (18.11.97)</td> <td>GB</td> </tr> <tr> <td>9725158.1</td> <td>27 November 1997 (27.11.97)</td> <td>GB</td> </tr> <tr> <td>9726147.3</td> <td>10 December 1997 (10.12.97)</td> <td>GB</td> </tr> <tr> <td>9800759.4</td> <td>14 January 1998 (14.01.98)</td> <td>GB</td> </tr> <tr> <td>9819016.8</td> <td>1 September 1998 (01.09.98)</td> <td>GB</td> </tr> </table>		9723516.2	6 November 1997 (06.11.97)	GB	9724190.5	14 November 1997 (14.11.97)	GB	9724386.9	18 November 1997 (18.11.97)	GB	9725158.1	27 November 1997 (27.11.97)	GB	9726147.3	10 December 1997 (10.12.97)	GB	9800759.4	14 January 1998 (14.01.98)	GB	9819016.8	1 September 1998 (01.09.98)	GB	(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
9723516.2	6 November 1997 (06.11.97)	GB																						
9724190.5	14 November 1997 (14.11.97)	GB																						
9724386.9	18 November 1997 (18.11.97)	GB																						
9725158.1	27 November 1997 (27.11.97)	GB																						
9726147.3	10 December 1997 (10.12.97)	GB																						
9800759.4	14 January 1998 (14.01.98)	GB																						
9819016.8	1 September 1998 (01.09.98)	GB																						
(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).		Published With international search report.																						
(72) Inventors; and (75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).		(88) Date of publication of the international search report: 2 March 2000 (02.03.00)																						
(54) Title: NEISSERIAL ANTIGENS																								
(57) Abstract The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.																								

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 98/01665

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/31 C07K14/22 C07K16/12 G01N33/53 A61K39/095
 //(C12N15/31,C12R1:36)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 99 55873 A (SMITHKLINE BEECHAM BIOLOGICALS S.A.) 4 November 1999 (1999-11-04) SEQ ID NOS: 1, 2, 3 and 4 page 82 -page 96 claims 3-5,10,11,17-21 --- -/-	5-8, 10-16



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

Z document member of the same patent family

Date of the actual completion of the international search

1 December 1999

Date of mailing of the international search report

08 December 1999 (08.12.99)

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

Fuchs, U

INTERNATIONAL SEARCH REPORT

International Application No

PCT, IB 98/01665

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>ROKBI, B. ET AL.: "Evaluation of Recombinant Transferrin-Binding Protein B Variants from <i>Neisseria meningitidis</i> for Their Ability To Induce Cross-Reactive and Bactericidal Antibodies against a Genetically Diverse Collection of Serogroup B Strains"</p> <p>INFECTION AND IMMUNITY, vol. 65, no. 1, January 1997 (1997-01), pages 55-63, XP002086937 abstract</p> <p>page 55, column 1, line 1 -page 56, column 2, line 78</p> <p>page 57, column 1, line 31 -page 61, column 2, line 63</p> <p>page 59; figure 3; table 2</p> <p>---</p>	1-17
A	<p>WO 96 12020 A (OREGON HEALTH SCIENCES UNIVERSITY) 25 April 1996 (1996-04-25) abstract</p> <p>page 32 -page 38; examples 2-4</p> <p>page 42 -page 44; example 9</p> <p>page 45; table 3</p> <p>---</p>	1-17
A	<p>WO 96 31618 A (THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL) 10 October 1996 (1996-10-10) abstract</p> <p>page 19, line 21 -page 20, line 13</p> <p>page 20, line 22 -page 21, line 11</p> <p>page 22, line 23 -page 24, line 16</p> <p>-----</p>	1-17

INTERNATIONAL SEARCH REPORT

Int. application No.
PCT/IB 98/01665

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

((1-3) completely) and ((4-17) partially)

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1. Claims: ((1-3) completely) and ((4-17) partially)

A protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2,4,6 and 8 or fragments thereof; a protein having 50% or greater sequence identity to said protein; an antibody binding said protein; a nucleic acid encoding said protein; a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOS: 1,3,5 and 7 or fragments thereof; a composition comprising said protein, said nucleic acid or said antibody; the use of said composition;

Invention 2 to 104. Claims (4-17) partially

Idem as subject 1 but limited to the ORFs corresponding to examples 2-104 characterized by SEQ ID NOS: 9-892.

(Invention 2 is limited to SEQ ID NOS: 9-10; Invention 3 is limited to SEQ ID NOS: 11-18; Invention 4 is limited to SEQ ID NOS: 19-28;; Invention 104 is limited to SEQ ID NOS: 885-892).

In view of additional search fees paid, Inventions 5, 26, 55, 77 and 91 have been further searched.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT, IB 98/01665

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9955873	A	04-11-1999	NONE	
WO 9612020	A	25-04-1996	US 5698438 A	16-12-1997
			AU 705509 B	27-05-1999
			AU 4007395 A	06-05-1996
			CA 2203116 A	25-04-1996
			EP 0793720 A	10-09-1997
			FI 971634 A	16-06-1997
			HU 77048 A	02-03-1998
			JP 10508469 T	25-08-1998
			NO 971768 A	03-06-1997
WO 9631618	A	10-10-1996	AU 5537096 A	23-10-1996
			CA 2217522 A	10-10-1996
			EP 0830456 A	25-03-1998
			JP 11503322 T	26-03-1999

